(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 July 2001 (19.07.2001)

PCT

(10) International Publication Number
WO 01/51659 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/IB01/00116

(22) International Filing Date: 11 January 2001 (11.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/175,854 13 January 2000 (13.01.2000) US

(71) Applicant *(for all designated States except US)*: **GENSET** [FR/FR]; Intellectual Property Dept., 24, rue Royale, F-75008 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHU, Tom** [US/US]; 5394 Camino Playa Norte, San Diego, CA 92124 (US). **BLUMENFELD, Marta** [FR/FR]; 5, rue Tagore, F-75013 Paris (FR). **COHEN, Daniel** [FR/FR]; 1, boulevard Richard Wallace, F-92200 Neuilly-sur-Seine (FR).

(74) Common Representative: **GENSET**; Intellectual Property Dept., 24, rue Royale, F-75008 Paris (FR).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/51659 A2

(54) Title: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING GENES INVOLVED IN CENTRAL NERVOUS SYSTEM DISORDERS

(57) Abstract: The invention provides polynucleotides including biallelic markers derived from genes involved in CNS disorders and from genomic regions flanking those genes. Primers hybridizing to regions flanking these biallelic markers are also provided. This invention also provides polynucleotides and methods suitable for genotyping a nucleic acid containing sample for one or more biallelic markers of the invention. Further, the invention provides methods to detect a statistical correlation between a biallelic marker allele and a phenotype and/or between a biallelic marker haplotype and a phenotype.

BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING GENES INVOLVED IN CENTRAL NERVOUS SYSTEM DISORDERS

FIELD OF THE INVENTION

The present invention is in the field of pharmacogenomics, and is primarily directed to biallelic markers that are located in or in the vicinity of genes that play a role in disorders of the brain and nervous system and to the uses of these markers. The present invention encompasses methods of establishing associations between these markers and central nervous system (CNS) disorders such as psychiatric disorders and neurodegenerative diseases as well as associations between these markers and treatment response to a variety of therapeutic agents. The present invention also provides means to determine the genetic predisposition of individuals to such diseases and means to predict responses to such drugs.

BACKGROUND OF THE INVENTION

Advances in the technological armamentarium available to basic and clinical investigators have enabled increasingly sophisticated studies of brain and nervous system function in health and disease. Numerous hypotheses both neurobiological and pharmacological have been advanced with respect to the neurochemical and genetic mechanisms involved in central nervous system (CNS) disorders, including psychiatric disorders and neurodegenerative diseases. However, CNS disorders have complex and poorly understood etiologies, as well as symptoms that are overlapping, poorly characterized, and difficult to measure. As a result future treatment regimes and drug development efforts will be required to be more sophisticated and focused on multigenic causes, and will need new assays to segment disease populations, and provide more accurate diagnostic and prognostic information on patients suffering from CNS disorders.

A. Neurological Basis of CNS Disorders

Neurotransmitters serve as signal transmitters throughout the body; therefore, diseases that affect neurotransmission can have serious consequences. For example, for over 30 years the leading theory to explain the biological basis of many psychiatric disorders such as depression has been the monoamine hypothesis. This hypothesis proposes that depression is partially due to a deficiency in one of the three major biogenic monoamines, namely dopamine, norepinephrine and serotonin. However, this hypothesis has been replaced by one that takes into account the overall function of the brain and no longer only considers a single neuronal system.

Dopamine

Dopamine is synthesized via the hydroxylation of tyrosine to dihydroxyphenylalanine (DOPA), involving the enzyme tyrosine hydroxylase (TH) and catechol O-methyl transferase (see Figure 2: TH and COMT). Tyrosine hydroxylase is the rate-limiting enzyme in the synthesis

of catecholamines such as dopamine and norepinephrine. Dopamine and the enzymes involved in its biosynthesis and degradation are known to be involved in the pathophysiology of a depression, schizophrenia and Parkinson's disease. For example, it is believed tyrosine hydroxylase may be involved in the pathophysiology of psychiatric disorders and positive associations have been reported for tyrosine hydroxylase gene markers in mood disorders. However, a recent study was unable to conclude tyrosine hydroxylase variants are related with depressive symptomatology in subjects affected by mood disorder (Serretti A. et al.; *American Journal of Medical Genetics* 81(2):127-130, 1998).

Dopamine, released from the nerve terminals, is largely recaptured by a re-uptake mechanism involving dopamine transporter (DAT) (see Table 2: DAT). Following re-uptake, dopamine is metabolized by monoamine oxidases A and B (MAOA/B) (see Table 2: MAOA and MAOB). Monoamine oxidase A and B are critical enzymes in deamination of biogenic amines and may be involved in the pathophysiology of major psychoses, including mood disorder, Parkinson's disease and schizophrenia. Recently, evidence for genetic association between the MAOA gene and bipolar mood disorder was demonstrated in a Caucasian population, but not seen in a Japanese population (Sasaki T. et al.; *Biological Psychiatry* 44(9):922-924, 1998).

Receptors for dopamine regulate dopaminergic neurotransmission. A plethora of dopamine receptors exist, including the presynaptic dopamine transporter and at least five pharmacologic subtypes (D_1 -which is linked to the enzyme adenyl cyclase, D_2 -not linked to adenyl cyclase, D_3, D_4, and D_5). Classically, the most extensively investigated dopamine receptor is the D_2 receptor, as it is stimulated by dopaminergic agonists for the treatment of Parkinson's disease and blocked by dopamine antagonist neuroleptics for the treatment of schizophrenia (see Table 2: DRD2). Recently, other dopamine receptors, particularly the D_4 receptor, have become targets for new antipsychotics in the treatment of Parkinson's disease (see Table 2: DRD4). It appears the interaction of all or some of the dopamine receptors play a role in many CNS disorders. However, only a limited number of studies investigating the association of such disorders with genes of the dopaminergic pathway have been completed and often with conflicting results.

Norepinephrine

The noradrenergic system is known to play a large role in the determination of mood, dysfunction of contributes to the "functional" disorders of depression, mania and anxiety. It is believed depressed patients are unable to produce sufficient norepinephrine in some parts of the brain for neuronal transmission, while mania may result from excessive activity or sensitivity of this system.

The amino acid tyrosine, having been actively taken up by adrenergic neurons, is converted to DOPA by means of tyrosine hydroxylase. DOPA is then converted to dopamine and later into norepinephrine in the synaptic vesicles. Conversion of norepinephrine to

epinephrine occurs in the adrenal medulla and also in certain restricted parts of the brain. Catecholamine degradation is enzymatically controlled by MAOA/B intraneuronally, with the main norepinephrine metabolite being 3-methoxy-4-hydroxyphenylglycol.

The noradrenergic neuron is regulated by a multiplicity of receptors and for norepinephrine, these being designated α_1, α_2, β_1 and β_2. Postsynaptic norepinephrine receptors bind norepinephrine released from the presynaptic neuron and activate a molecular cascade in the postsynaptic neuron. Specifically, the activation of α_2 receptors causes inhibition of norepinephrine, whereas activation of β1 receptors leads to increased release of norepinephrine from adrenergic terminals (see Table 2: ADRB1R). Systemically, the adrenoreceptor subtypes α1, α2, β1 and β2 are functional in a variety of other ways ranging from vasodilatation to initiating smooth muscle relaxation. The action of norepinephrine is terminated by the norepinephrine transporter (NET), a membrane protein that serves as a reuptake pump for synaptic norepinephrine (see Table 2: NET). These receptors and transporter are the target of many therapeutic agents currently used to treat psychiatric disorders particularly depression.

Serotonin (5-hydroxytryptamine, 5HT)

The serotoninergic system is an anatomically diverse system with pathways that follow closely those of the noradrenergic system, but are quite different from those of the dopaminergic distribution. The physiological functions in which the serotoninergic system is involved include sleep, appetite, nociception, diurnal rhythmicity, neuroendocrine regulation and mood. At the level of consciousness there is also the suggestion that rational thought processes arise, using previously stored information, with the aid of the serotoninergic system. Serotoninergic projections innervating the hypothalamus influence the secretion of several anterior pituitary hormones. There is evidence that serotonin may serve as the final common pathway by which other neurotransmitters act in controlling secretion of many hormones.

Tryptophan is taken up by active transport into the neurons where it is hydroxylated by tryptophan hydroxylase to 5-hydroxytryptophan (5HTP). The latter is then decarboxylated to serotonin which, following release from the neurons, is recovered by a re-uptake mechanism. Degradation of serotonin occurs by way of MAOA/B and the majority of the metabolites are excreted in the urine.

Serotonin receptors come in 13 or more subtypes that can vary in their sensitivity to serotonin and in the effects they produce. An increasingly complex series of serotonin receptors is being identified. Presynaptic serotonin uptake sites and serotonin receptors designated $5HT_1$ (and further subdivided into $5HT_{1a}$, $5HT_{1b}$, $5HT_{1c}$), $5HT_2$, $5HT_3$, $5HT_4$, and $5HT_6$, have been identified by means of pharmacological studies (see Table 2: 5HTT, 5HT1A, 5HTR2C, 5HTR6, and 5HTR7). As a whole, communication between two neurons is complex and may be

mediated by more than one neurotransmitter; for example, the serotonergic system may co-exist with other neurotransmitter in the same synapses.

Gamma aminobutyric acid (GABA)

Gamma aminobutyric acid (GABA) is an important amino acid which functions as the most prevalent inhibitory neurotransmitter in the central nervous system. Gamma aminobutyric acid works in partnership with a derivative of Vitamin B-6, pyridoxine, to cross from the axons to the dendrites through the synaptic cleft, in response to an electrical signal in the neuron and inhibits message transmission. This helps control the nerve cells from firing too fast, which would overload the system.

The gamma aminobutyric acid (a) receptor (see Table 2: GABRA5 and GABRG2) appears to play a key role in modulating anxiety and could be involved in either the etiology or the pathogenesis of anxiety disorders (Crowe et al. *Am J Psychiatry* 154:8). A benzodiazepine binding site is located on this receptor, and ligands that bind to this site can either increase or decrease anxiety.

Growth associated protein (GAP43)

Growth associated protein (GAP43) is localized exclusively to nerve tissue and is known to play a role in synaptic transmission and membrane permeability. The expression of GAP43 is associated with mammalian peripheral nerve regeneration (Kosik et al. *Neuron* 1:127-132, 1988). A polymorphism in the 3'-untranslated region of GAP43 is found at slightly lower frequencies in Alzheimers and Parkinson's patients (Poduslo. *Hum Genet.* 92:635-636, 1993).

Subreceptor Activity

The activity of subreceptors has also been investigated in recent years for its role in a wide range of CNS disorders. When neurotransmitters bind to receptors on the membranes of postsynaptic neurons, they elicit a target response in the cell via a second or third messenger. Several different messenger chemicals are known including cyclic adenosine monophosphate (AMP). G proteins serve as signal transduction subunits in the cyclic AMP pathway (see Table 2: Gbeta3). G protein coupled receptors are thought to have seven membrane spanning domains and have been divided into 2 subclasses: those in which the binding site is in the extracellular domain for example receptors for glycoprotein hormones, such as thyroid stimulating hormone (TSH) and follicle stimulating hormone (FSH) and those in which the ligand binding site is likely to be in the plane of the 7 transmembrane domains for example rhodopsin and receptors for small neurotransmitters and hormones for example muscarinic acetylcholine receptor. However, orphan G-protein coupled receptor (see Table 2: HM77) does not contain N-linked glycosylation sites near the N-terminus like other members of this protein family.

There is evidence that various monoamine or monoaminergic receptors are able to alter cyclic AMP levels through the same G protein. Therefore, G proteins serve as an important cross-talk mechanism between transmitter systems in the CNS. An abnormality in a G protein or

in the responsiveness of cyclic AMP to any of the receptors may result in an alteration in monoaminergic neurotransmission. For example, guanine nucleotide binding protein olfactory type (see Table 2: GOLF) is believed to play a role in signaling involving the cAMP mediated signaling pathway as well as the norepinephrine pathway.

B. Endocrine Basis of CNS Disorders

Biological theories of many CNS disorders have long revolved around the main monoamine systems, namely dopamine, norepinephrine and serotonin. It is apparent, however, that these three systems do not completely explain the pathophysiology of many CNS disorders. The hypothalamic-pituitary-adrenal (HPA) axis, including the effects of corticotrophin-releasing factor and glucocorticoids, plays an important role in the pathophysiology of CNS disorders.

The hypothalamus lies at the top of the hierarchy regulating hormone secretion via the hypothalamus-pituitary-adrenal (HPA) axis. It manufactures and releases peptides that act on the pituitary, thus stimulating or inhibiting the pituitary's release of various hormones into the blood. These hormones, among them growth hormone, thyroid-stimulating hormone and adrenocorticotrophic hormone (ACTH), control the release of other hormones from target glands. In addition to functioning outside the nervous system, the hormones released in response to pituitary hormones feed back to the pituitary and hypothalamus. There they deliver inhibitory signals that serve to limit excess hormone biosynthesis.

Also included in the regulation of the HPA axis is vasopressin receptor 1A (see Table 2: AVPR1A). Vasopressin receptors are present in a number of tissues including the anterior pituitary, where they stimulate adrenocorticotrophic hormone (ACTH) release (Thibonnier et al. *Genomics* 31: 327-334, 1996).

Dysregulation of the HPA axis appears to be an important feature of many psychiatric disorders and neurodegenerative diseases. When a threat to physical or psychological well-being is detected, the hypothalamus amplifies production of corticotrophin-releasing factor (CRF), which induces the pituitary to secrete ACTH (see Table 2: CRF, CRHBP, CRFR1 and CRFR2). ACTH then instructs the adrenal glands to release cortisol. Therefore, it is believed chronic activation of the HPA axis may lay the ground for illness.

The increased HPA drive is primarily mediated by hypersecretion of corticotrophin-releasing factor. Patients with major depression show increased levels of lumbar cerebrospinal fluid (CSF) corticotrophin-releasing factor as compared to matched controls or patients with other neurologic illnesses (Plotsky, P.M., *Psych. Clin. Of North Am.*, 21(2):293-307, 1998). Dysregulation of hypothalamic corticotrophin-releasing factor neurons, whether intrinsic or extrinsic to these neurons, can result in corticotrophin-releasing factor hypersectretion leading to elevations in cortisol followed by adaptive down regulation of both pituitary and central glucocorticoid receptors and corticotrophic-releasing receptors.

The anxiolytic effects of corticotrophin-releasing factor appear to be mediated by the activation of the central noradrenergic system. A CRF-positive projection has been identified linking limbic structures to the noradrenergic locus ceruleus, stimulation of which plays an important role in emotional memory and increases tyrosine hydroxylase activity. Therefore, primary or secondary dysfunction of corticotrophin-releasing factor would be expected to initiate a cascade of maladaptations.

Glucocorticoids and mineralocorticoids are both classes of steroid hormones that play an important role in the HPA axis; and have therefore been implicated in the pathophysiology of various psychiatric disorders and neurodegenerative diseases (see Table 2: GRL and MLR). Glucocorticoids exert numerous effects on metabolism, reproduction, inflammation and immunity. In addition, glucocorticoids serve as the primary negative feedback mechanism that regulates the HPA axis. Mineralocorticoids maintain electrolyte balance by regulating salt and water retention in the kidneys.

Brain-derived neurotrophic factor (see Table 2: BDNF) is a member of a group of proteins that includes neurotrophin-3/4/5 and nerve growth factor (NGF) and are believed to play a role in the etiology of a number of CNS-related disorders including schizophrenia and Parkinson's disease (Hawi et al. *Psychiatry Research* 81: 111-116, 1998 and Gasser et al. *Annals of Neurology* 36(3)387-396, 1994). BDNF plays an important role in promoting growth and maintenance during normal development and differentiation of the vertebrate system (Hanson et al. *Genomics* 13: 1331-1333, 1992). Further, it is believed BDNF has an effect on the differentiation of dopaminergic and serotonergic neurons (Studer et al. *Euro. J. of Neuroscience* 7: 223-233, 1995).

C. Examples of CNS Disorders

Neurotransmitter and hormonal abnormalities are implicated in disorders of movement (e.g. Parkinson's disease, Huntington's disease, motor neuron disease, etc.), disorders of mood (e.g. unipolar depression, bipolar disorder, anxiety, etc.) and diseases involving the intellect (e.g. Alzheimer's disease, Lewy body dementia, schizophrenia, etc.). In addition, Neurotransmitter and hormonal abnormalities have been implicated in a wide range of disorders, such as coma, head injury, cerebral infarction, epilepsy, alcoholism and the mental retardation states of metabolic origin seen particularly in childhood.

Schizophrenia

In developed countries schizophrenia occurs in approximately one per cent of the adult population at some point during their lives. There are an estimated 45 million people with schizophrenia in the world, with more than 33 million of them in the developing countries. Moreover, schizophrenia accounts for a fourth of all mental health costs and takes up one in three psychiatric hospital beds. Most schizophrenia patients are never able to work. The cost of schizophrenia to society is enormous. In the United States, for example, the direct cost of

treatment of schizophrenia has been estimated to be close to 0.5% of the gross national product. Standardized mortality ratios (SMRs) for schizophrenic patients are estimated to be two to four times higher than the general population and their life expectancy overall is 20 % shorter than for the general population.

The most common cause of death among schizophrenic patients is suicide (in 10% of patients) which represents a 20 times higher risk than for the general population. Deaths from heart disease and from diseases of the respiratory and digestive system are also increased among schizophrenic patients.

Schizophrenia comprises a group of psychoses with either 'positive' or 'negative' symptoms. Positive symptoms consist of hallucinations, delusions and disorders of thought; negative symptoms include emotional flattening, lack of volition and a decrease in motor activity.

A number of biochemical abnormalities have been identified and, in consequence, several neurotransmitter-based hypotheses have been advanced over recent years; the most popular one has been "the dopamine hypothesis," one variant of which states that there is over-activity of the mesolimbic dopamine pathways at the level of the D_2 receptor. However, researchers have been unable to consistently find an association between various receptors of the dopaminergic system and schizophrenia.

In addition to the hypotheses which are briefly presented here, and which attempt to draw together the neurochemical observations in schizophrenia, one should add that some abnormalities of cortical neuropeptides are well documented. These abnormalities include changes in the levels of somatostatin, substance P, cholecystokinin (CCK) and vasoactive intestinal peptide (VIP) found in association with negative symptom defect states in the temporal area of the brain (i.e. the hippocampus, amygdala and neocortex), in particular.

Bipolar Disorder

Bipolar disorders are relatively common, occurring in about 1.3% of the population, and have been reported to constitute about half of the mood disorders seen in psychiatric clinics. Bipolar disorders have been found to vary with gender depending of the type of disorder; for example, bipolar disorder I is found equally among men and women, while bipolar disorder II is reportedly more common in women. The age of onset of bipolar disorders is typically in the teenage years and diagnosis is typically made in the patient's early twenties. Bipolar disorders also occur among the elderly, generally as a result of a neurological disorder or other medical conditions. In addition to the severe effects on patients' social development, suicide completion rates among bipolar patients are reported to be about 15%.

Bipolar disorders are characterized by phases of excitement and depression; the excitement phases (mania) and depressive phases can alternate or occur in numerous admixtures with varying degrees of severity and duration. Because bipolar disorders can exist in different

forms and display different symptoms, the classification of bipolar disorder has been the subject of extensive studies resulting in the definition of bipolar disorder subtypes and widening of the overall concept to include patients previously thought to be suffering from different disorders. Bipolar disorders often share certain clinical signs, symptoms, treatments and neurobiological features with psychotic illnesses in general and therefore present a challenge to the psychiatrist to make an accurate diagnosis. Furthermore, because the course of bipolar disorders and various mood and psychotic disorders can differ greatly, it is critical to characterize the illness as early as possible in order to offer means to manage the illness over a long term.

Diagnosis of bipolar disorder can be very challenging. One particularly troublesome difficulty is that some patients exhibit mixed states, simultaneously manic and dysphoric or depressive, but do not fall into the DSM-IV classification because not all required criteria for mania and major depression are met daily for at least one week. Other difficulties include classification of patients in the DSM-IV groups based on duration of phase since patients often cycle between excited and depressive episodes at different rates. In particular, it is reported that the use of antidepressants may alter the course of the disease for the worse by causing "rapid-cycling". Also making diagnosis more difficult is the fact that bipolar patients, particularly at what is known as Stage III mania, share symptoms of disorganized thinking and behavior with bipolar disorder patients. Furthermore, psychiatrists must distinguish between agitated depression and mixed mania; it is common that patients with major depression exhibit agitation, resulting in bipolar-like features.

For both schizophrenia and bipolar disorder, all the known molecules used for treatment have side effects and act only against the symptoms of the disease. There is a strong need for new molecules without associated side effects or reduced side effects which are directed against targets that are involved in the causal mechanisms of schizophrenia and bipolar disorder. Therefore, tools facilitating the discovery and characterization of these targets are necessary and useful.

The aggregation of schizophrenia and bipolar disorder in families, the evidence from twin and adoption studies, and the lack of variation in incidence worldwide, indicate that schizophrenia and bipolar disorder are primarily genetic conditions, although environmental risk factors are also involved at some level as necessary, sufficient, or interactive causes. For example, schizophrenia occurs in 1% of the general population. However, if a subject has one grandparent with schizophrenia, the risk of getting the illness increases to about 3%, while one parent with Schizophrenia increases risk to about 10%. When both parents have schizophrenia, the risk rises to approximately 40%. Consequently, there is a strong need to identify genes involved in schizophrenia and bipolar disorder. The knowledge of these genes will allow researchers to understand the etiology of schizophrenia and bipolar disorder and could lead to

drugs and medications which are directed against the cause of the diseases, not just against their symptoms.

There is also a great need for new methods to detect susceptibility to schizophrenia and bipolar disorder, as well as for preventing or following up the development of the disease. Diagnostic tools could also prove extremely useful. Indeed, early identification of subjects at risk of developing schizophrenia would enable early and/or prophylactic treatment to be administered. Moreover, accurate assessments of the eventual efficacy of a medicament as well as the patent's eventual tolerance to it may enable clinicians to enhance the benefit/risk ratio of schizophrenia and bipolar disorder treatment regimes.

Depression

Depression is a serious medical illness that affects 340 million people worldwide. In contrast to the normal emotional experiences of sadness, loss, or passing mood states, clinical depression is persistent and can interfere significantly with an individual's ability to function. As a result, depression is the leading cause of disability throughout the world with an estimated cost of $53 billion each year in the United States alone.

Symptoms of depression include depressed mood, diminished interest or pleasure in activities, change in appetite or weight, insomnia or hypersomnia, psycho-motor agitation or retardation, fatigue or loss of energy, feelings of worthlessness or excessive guilt, anxiety, inability to concentrate or act decisively, and recurrent thoughts of death or suicide. A diagnosis of unipolar major depression (or major depressive disorder) is made if a person has five or more of these symptoms and impairment in usual functioning nearly every day during the same two-week period. The onset of depression generally begins in late adolescence or early adult life; however, recent evidence suggests depression may be occurring earlier in life in people born in the past thirty years.

The World Health Organization predicts that by the year 2020 depression will be the greatest burden of ill-health to people in the developing world, and that by then depression will be the second largest cause of death and disability. Beyond the almost unbearable misery it causes, the big risk in major depression is suicide. Within five years of suffering a major depression, an estimated 25% of sufferers try to kill themselves. In addition, depression is a frequent and serious complication of heart attack, stroke, diabetes, and cancer. According to one recent study that covered a 13-year period, individuals with a history of major depression were four times as likely to suffer a heart attack compared to people without such a history.

Depression may be a feature in up to 50% of patients with CNS disorders such as Parkinson's disease and Alzheimer's disease. The neuronal loss in the locus ceruleus, typical of Alzheimer's disease, is greatest in those patients who have depression; such patients also have lower norepinephrine levels than do those who lack depressive features. Approximately 50% of

patients with Alzheimer's disease have less norepinephrine than normal in the majority of cortical and subcortical areas of the brain that have been examined to date.

Many neurochemical findings are coming to light implicating a biological basis for the depression, at least for certain subtypes. Abnormalities of monoamine function have been recognized in depression for many years involving norepinephrine, serotonin and dopamine. Changes in adrenoceptor density and function as well as changes in adrenoceptors associated with the pituitary-adrenal axis function strongly implicate a disorder in central noradrenergic transmission in depression. This dysfunction may be caused by changes in the activity of tyrosine hydroxylase. The effect of corticotrophin releasing factor in modulating the activity of noradrenergic neurons in the locus ceruleus may provide the link between environmental trigger factors and central noradrenergic dysfunction, along with dysfunction of the HPA axis.

Dysfunction of serotonin metabolism, as shown by decreased concentrations of the metabolite 5HIAA in cerebrospinal fluid (CSF), is linked with depression; nevertheless, it is not a feature in all patients with depression. Therefore, a subgroup entitled "serotonin depression" has been proposed. Often included among those who suffer from serotonin depression are patients who also suffer a number of neurological diseases. A reduction in the number of serotonin-containing neurons in the median raphe in Parkinson's disease, Alzheimer's disease and, possibly, the elderly, is associated with the development of depression.

Low levels of the dopamine metabolite HVA are found in the CSF in patients with depression. In addition, dopamine agonists produce a therapeutic response in depression.

Presently, antidepressants are designed to address many of the symptoms of depression by increasing neurotransmitter concentration in aminergic synapses. Distinct pharmacologic mechanisms allow the antidepressants to be separated into seven different classes. The two classical mechanisms are those of tricyclic antidepressants (TCAs) and monoamine oxidase inhibitors (MAOIs). The most widely prescribed agents are the serotonin selective reuptake inhibitors (SSRIs). Three other classes of antidepressants, like the SSRIs, increase serotonergic neurotransmission, but they also have additional actions, namely dual serotonin and norepinephrine reuptake inhibition; serotonin-2 antagonism/reuptake inhibition; and α_2 antagonism plus serotonin-2 and -3 antagonism. The selective norepinephrine and dopamine reuptake inhibitors define a novel class of antidepressant that has no direct actions on the serotonin system.

Recent findings suggest some re-appraisal and modifications of the monoamine hypothesis are necessary. The increased levels of monoamine transmitters at the synapses, although quickly produced in response to antidepressant therapy, are in contrast with the much slower clinical recovery of the patient from depression, which takes about two weeks to begin and may only reach maximal levels several weeks later. Moreover, should acute depletion of either norepinephrine and/or serotonin occur experimentally in a normal individual, then

depression does not, in the short-term, occur. Not in keeping with the hypothesis, too, is the cerebral resistance generated in response to the pharmacological changes induced by antidepressant compounds. These counteractive changes comprise reduction in the number of post-synaptic β-receptors, together with a lowered firing rate of noradrenergic neurons.

In addition, there are subsets of patients with differential responses to antidepressants. Thus far, biochemical predictors of treatment response have failed to identify definite parameters that could correctly identify patients more likely to respond to particular classes of antidepressants (Schatzberg, Alan F., *Journal of Clinical Psychiatry*, 59:15-18, 1998). As a result, psychiatrists often must choose a treatment based on intuition or trail and error. However, probes such as biallelic markers could serve as an invaluable tool to successfully identify patients who might respond preferentially to existing and new antidepressants (Charney, Dennis S, *Journal of Clinical Psychiatry*, 59:11-14, 1998). In particular, markers from genes known to affect drug response such as transcription factors (see Table 2: SEF-1B) and drug metabolizing enzymes (see Table 2: CYP3A4) need to be investigated to determine "responders" and "non-responders" to medicaments.

While modulating monoamine activity as a therapeutic strategy continues to dominate research, an important new development has been the emergence of novel mechanisms of action, notably modulation of the activity of neuropeptides, namely through the neuropeptide receptor Y1, the tachykinin NK1 receptor and nicotinic receptors (see Table 2: NPY1R, TACR1 and CHRNA7). Recent clinical trails showed that tachykinin NK1 receptor antagonists are effective in treating depression and chemotherapy-induced emesis. Therefore, it is well possible that such antagonists will be clinically useful for treatment of specific CNS disorders. Nicotinic receptors are known to serve as important ligand-gated ion channels active in classical, excitatory neurotransmission and perhaps more novel forms of neurochemical signaling. Their critical functional roles both centrally and peripherally make them ideal targets for regulation of the nervous system. Finally, new antidepressants that may render the HPA axis more sensitive to glucocorticoid feedback are being investigated as well.

In addition to monoamine dysfunction as a possible cause of depression, researchers have reported increased activity in the HPA axis in untreated depressed patients, as evinced by raised levels of cortisol in urine, blood and cerebrospinal fluid, as well as by other measures. Numerous studies have confirmed that substantial numbers of depressed patients, particularly those most severely affected, display HPA axis hyperactivity. Patients with depression frequently have symptom clusters which point strongly to involvement of the HPA system as a relay station between neurocircuitries in the brain and peripheral hormone and autonomic nervous function. It has been proposed that this increased, state-dependent hyperactivity of the HPA system in depression is probably initiated and/or maintained by the combination of enhanced central production of corticotrophin-releasing factor and desensitization of the binary,

glucocorticoid receptor binding system in the hippocampus, which is the central regulator of HPA system activity.

Deeper investigation of the phenomenon has now revealed alterations at each level of the HPA axis in depressed patients. For instance, both the adrenal gland and the pituitary are enlarged, and the adrenal gland hypersecretes cortisol. But many researchers, have become persuaded that aberrations in CRF-producing neurons of the hypothalamus and elsewhere bear most of the responsibility for HPA axis hyperactivity and the emergence of depressive symptoms.

Many studies have shown corticotrophin-releasing factor concentrations in cerebrospinal fluid to be elevated in depressed patients, compared with control subjects or individuals with other psychiatric disorders. This magnification of corticotrophin-releasing factor levels is reduced by treatment with antidepressants and by effective electroconvulsive therapy. Further, postmortem brain tissue studies have revealed a marked exaggeration both in the number of CRF-producing neurons in the hypothalamus and in the expression of the corticotrophin-releasing factor gene (resulting in elevated corticotrophin-releasing factor synthesis) in depressed patients as compared with controls. Moreover, delivery of corticotrophin-releasing factor to the brains of laboratory animals produces behavioral effects that are cardinal features of depression in humans, namely, insomnia, decreased appetite, decreased libido and anxiety.

Geneticists have provided some of the oldest proof of a biological component to depression in many people. Depression and manic-depression frequently run in families. Thus, close blood relatives of patients with severe depressive or bipolar disorder are much more likely to suffer from those or related conditions than are members of the general population. Studies of identical and fraternal twins also support an inherited component. Illness in both members of a pair is much higher for manic-depression in identical twins than in fraternal and is somewhat elevated for depression alone.

In the past 20 years, genetic researchers have expended great effort trying to identify the genes which contribute to depression. So far, though, those genes have evaded discovery, perhaps because a predisposition to depression involves several genes, each of which makes only a small, hard-to-detect contribution. As a result, psychiatrists today have to choose antidepressant medications by intuition and trial and error; a situation that can put suicidal patients in jeopardy for weeks or months until the right compound is selected. Therefore, there is a strong need to successfully identify genes involved in depression; thus allowing researchers to understand the etiology of depression and address its cause, rather than symptoms.

Alzheimer's Disease

Alzheimer's disease is characterized by the onset in middle age of a slowly progressive dementia; there is loss of memory for past events, inability to develop new memories and impairment of intellect, all leading to a lessened capacity for dealing with the tasks and problems

of daily living. It is the most common cause of both presenile and senile dementia. Alzheimer's disease is not the non-specific degenerative disorder of the CNS that it was once thought to be, as neurochemical studies on postmortem material now reveal the degeneration to be selective for certain neuronal populations in the subcortical and cortical areas; other cell populations seem to be unaffected. Senile plaques and neurofibrillary tangles are the characteristic histological feature, found throughout the cerebral cortex and especially in certain regions of the limbic system (the amygdala and hippocampus), perhaps accounting for the memory loss so typical of the early phase of the disease. In addition, there is reduction of acetylcholine, norepinephrine, serotonin and somatostatin in the subcortical areas in Alzheimer's disease.

The activity of CAT, the enzyme involved in acetylcholine synthesis, is markedly decreased in Alzheimer's disease. This decrease does not occur in all areas of the brain, but does so particularly in the hippocampus and amygdala, which are some of the main sites where senile plaques and neurofibrillary tangles accumulate. The loss of such cortical cholinergic activity correlates well with the degree of dementia in patients with this disease. A further finding is that nerve growth factor (NGF) is now known to be involved in the maintenance of cholinergic neurons in the forebrain; also, nicotine, a cholinomimetic compound, is able to stimulate dopaminergic neurons via their nicotinic receptors; thus, seemingly, to provide smokers with some protection against degeneration of the dopaminergic neurons. The forebrain cholinergic system degenerates not only in Alzheimer's disease, but also in alcohol-induced dementia, Pick's disease, Lewy body dementia, progressive supranuclear palsy and in Parkinson's disease.

In Alzheimer's disease there is a reduction of both serotonin and its receptor proteins in the temporal lobe of the brain, as revealed from studies on autopsy and biopsy material. The loss of serotonin is, however, less than in Parkinson's disease and it would be unlikely, therefore, that the severe memory loss of Alzheimer's disease could be accounted for on this basis alone, although in Parkinson's disease there is an important difference in that the $5HT_2$ receptor is not decreased. Of interest in this context, but not necessarily related, is the bradyphrenia (characterized by difficulty in concentration, slowing of thought processes and inability to associate ideas) of Parkinson's disease where serotonin is low in most of the cortical regions. In the Lewy body type of senile dementia it is common for visual hallucinations to occur, and it is of great interest that in the temporal lobe the serotonergic activity is higher (as shown by the raised serotonergic:cholinergic ratio) in those patients who suffer from hallucinations compared with those who do not.

In addition to the involvement of serotonin in Alzheimer's disease, patients also suffer from decreased levels of norepinephrine and several neuropeptides. It is in those patients with Alzheimer's disease who also have depression that there is not only greatest reduction in the number of neurons within the locus ceruleus but also a markedly reduced norepinephrine content. There is also associated reduction in cortical somatostatin and corticotrophin-releasing factor,

and loss of the somatostatin content of neurons in the temporal cortex develops early in the condition.

There is no known definitive cure for Alzheimer's disease; therefore, treatment is aimed at relief of symptoms and protection from the effects of the deteriorating condition. Most treatments are still considered experimental or have had variable results. Treatment is also aimed at underlying disorders that contribute to confusion such as heart failure, hypoxia, thyroid disorders, anemia, nutritional disorders, infections, and psychiatric conditions such as depression. The correction of coexisting medical and psychiatric disorders often improves the patient's mental function.

Parkinson's Disease

Parkinson's disease is a disabling progressive neurodegenerative disorder characterized by tremor, rigidity, bradykinesia, and loss of postural reflexes. In the United States, about a million people are believed to suffer from Parkinson's disease, and about 50,000 new cases are reported every year. Because the symptoms typically appear later in life, these Tables are expected to grow as the average age of the population increases over the next several decades. The disorder is most frequent among people in there 70s and 80s, and appears to be slightly more common in men than in women. Parkinson's disease is found all over the world. The rates vary from country to country, but it is not clear whether this reflects true ethnic or geographic differences or simply variations in data collection.

The pathology is not completely understood, but there appears to be consistent changes in the melanin-containing nerve cells in the brainstem (substantia nigra, locus ceruleus), where there are varying degrees of nerve cell loss with reactive gliosis along with eosinophilic intracytoplasmic inclusions (Lewy bodies). As a result, the primary neurochemical defect in Parkinson's disease is the loss of dopaminergic projections to the striatum. Moreover, the loss of these populations of neurons also leads to neurotransmitter deficits, but to a lesser extent than that which accompanies the massive degeneration of dopaminergic neurons. For example, norepinephrine, serotonin and acetylcholine are variably decreased in Parkinson's disease due to loss of neurons in the locus ceruleus, raphe nuclei and the nucleus basalis of Meynert. Thus, some of the secondary clinical features of Parkinson's disease have been ascribed to these neurotransmitter deficits.

The neurochemical defect associated with Parkinson's disease can be partially corrected by L-DOPA, which helps replace the brain's dopamine, but cannot reverse the progression of the disease. There is no specific biological test for the diagnosis of Parkinson's disease. Twin studies have shown variable results and suggest that the genetics of this disorder will prove to be complex. Despite the importance and severity of Parkinson's disease and many years of research, a cause has not been identified and there is neither means of preventing the disease nor a proven permanent cure.

Findings of considerable importance in this search would be the location of a genetic marker, determination of the probability of penetrance, determination of possible genetic heterogeneity, and evidence of multifactorial inheritance with environmental interaction. Genetic factors determining susceptibility to Parkinson's disease will enhance epidemiological studies and possibly lead to identification of susceptible groups and of significant risk factors.

Huntington's Disease

Huntington's disease is a hereditary neurodegenerative disease that generally develops subtly in a person's thirties or forties; though it can begin any time between childhood and old age. In the United States alone, about 30,000 people have Huntington's disease, while at least 150,000 others have a 50 percent risk of developing the disease and thousands more of their relatives live with the possibility that they, too, might develop Huntington's disease.

Huntington's disease is characterized by difficulties in three areas: a movement disorder, dementia, and psychiatric disturbances. The movement disorder consists of two parts: involuntary twitching movement which first tend to involve the fingers and toes and then progress to include the whole body, and difficulties with voluntary movements in the form of clumsiness, stiffness, or trouble with walking. Dementia refers to a gradual loss of intellectual abilities such as memory, concentration, problem solving, and judgment. Psychiatric disturbances do not strike every person with Huntington's disease, but when they do, usually take the form of depression, irritability, and apathy. Depression and other psychiatric conditions in people with Huntington's disease, which seem to result from damage to the brain, can be debilitating.

Loss of neurotransmitter receptors, especially glutamate and dopamine receptors, is one of the pathologic hallmarks of patients with Huntington's disease (Cha J.H. et al.; *Proc National Acad Sci USA* may 26;95(11):6480-5, 1998). In addition, deficiency of GABA permits excessive dopaminergic activity in the corpus striatum resulting in onset of Huntington's disease, on account of the imbalance generated between cholinergic and dopaminergic systems.

Researchers have identified a single gene product thought to be causal when mutated by a tri-nucleotide repeat expansion. However, there is at present no cure for Huntington's disease or even any direct treatments, although researchers are presently working on a number of treatments which may slow down the progression of the disease. In the early and middle stages of the disease, medications called neuroleptics, which are given in larger doses for psychiatric complaints, can be given in small doses to Huntington's disease patients to suppress the involuntary movements. Drugs that cause increased dopamine release in the brain and dopamine receptor agonists are used, but both precipitate nausea and vomiting as side effects and dopamine antagonists are anti-emetic.

Pharmacogenomics and CNS Disorders

The vast majority of common diseases, such as all of the CNS disorders described above, are polygenic, meaning multiple genes cause them. In addition, these diseases are modulated by environmental factors such as pollutants, chemicals and diet. This is why many diseases are considered to be multifactorial; they result from a synergistic combination of factors, both genetic and environmental. Therapeutic management and drug development could be markedly improved by the identification of specific genetic polymorphisms that determine and predict patient susceptibility to diseases or patient responses to drugs.

To assess the origins of individual variations in disease susceptibility or drug response, pharmacogenomics uses the genomic technologies to identify polymorphisms within genes which are part of biological pathways involved in disease susceptibility, etiology, and development, or more specifically in drug response pathways responsible for a drug's efficacy, tolerance or toxicity. Pharmacogenomics can also provide tools to refine the design of drug development by decreasing the incidence of adverse events in drug tolerance studies, by better defining patient subpopulations of responders and non-responders in efficacy studies and, by combining the results obtained therefrom, to further allow better enlightened individualized drug usage based on efficacy/tolerance prognosis. Pharmacogenomics can also provide tools to identify new targets for designing drugs and to optimize the use of already existing drugs, in order to either increase their response rate and/or exclude non-responders from corresponding treatment, or decrease their undesirable side effects and/or exclude from corresponding treatment patients with marked susceptibility to undesirable side effects. However, for pharmacogenomics to become clinically useful on a large scale, additional molecular tools and diagnostics tests must become available.

Genetic Analysis of Complex Traits

Until recently, the identification of genes linked with detectable traits has relied mainly on a statistical approach called linkage analysis. Linkage analysis is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. Linkage analysis involves the study of families with multiple affected individuals and is useful in the detection of inherited-traits, which are caused by a single gene, or possibly a very small number of genes. Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns and which have a high penetrance (the probability that a person with a given genotype will exhibit a trait). About 100 pathological trait-causing genes have been discovered using linkage analysis over the last 10 years. But, linkage studies have proven difficult when applied to complex genetic traits. Most traits of medical relevance do not follow simple Mendelian monogenic inheritance. However, complex diseases often aggregate in families, which suggests that there is a genetic component to be found. Such complex traits are often due to the combined action of multiple genes as well as environmental factors. Such complex trait, include susceptibilities to heart disease, hypertension,

diabetes, cancer and inflammatory diseases. Drug efficacy, response and tolerance/toxicity can also be considered as multifactoral traits involving a genetic component in the same way as complex diseases. Linkage analysis cannot be applied to the study of such traits for which no large informative families are available. Moreover, because of their low penetrance, such complex traits do not segregate in a clear-cut Mendelian manner as they are passed from one generation to the next. Attempts to map such diseases have been plagued by inconclusive results, demonstrating the need for more sophisticated genetic tools.

Knowledge of genetic variation in the neuronal and endocrine systems is important for understanding why some people are more susceptible to disease or respond differently to treatments. Ways to identify genetic polymorphism and to analyze how they impact and predict disease susceptibility and response to treatment are needed.

Although the genes involved in the neuronal and endocrine systems represent major drug targets and are of high relevance to pharmaceutical research, we still have scant knowledge concerning the extent and nature of, sequence variation in these genes and their regulatory elements. In the case where polymorphisms have been identified the relevance of the variation is rarely understood. While polymorphisms hold promise for use as genetic markers in determining which genes contribute to multigenic or quantitative traits, suitable markers and suitable methods for exploiting those markers have not been found and brought to bare on the genes related to disorders of the brain and nervous system.

In the cases where polymorphisms have been identified, the relevance of the variation is rarely understood. While polymorphisms hold promise for use as genetic markers in determining which genes contribute to multigenic or quantitative traits, suitable markers and suitable methods for exploiting those markers have not been found and brought to bare on the genes related to central nervous system disorders.

SUMMARY OF THE INVENTION

The present invention is based on the discovery of a set of novel CNS disorder-related biallelic markers. See Table 7. These markers are located in the coding regions as well as non-coding regions adjacent to genes which express proteins associated with CNS disorders. The position of these markers and knowledge of the surrounding sequence has been used to design polynucleotide compositions which are useful in determining the identity of nucleotides at the marker position, as well as more complex association and haplotyping studies which are useful in determining the genetic basis for disease states involving the neuronal and endocrine systems. In addition, the compositions and methods of the invention find use in the identification of the targets for the development of pharmaceutical agents and diagnostic methods, as well as the characterization of the differential efficacious responses to and side effects from pharmaceutical agents acting on CNS disorders. Further, the compositions and methods of the invention may be

employed in a process for screening for antagonists and/or agonists for the polypeptides of the invention. Such molecules may prove useful as therapeutics in the diagnosis and/or treatment of CNS disorders, particularly depression.

A first embodiment of the invention encompasses polynucleotides consisting of, consisting essentially of, or comprising a contiguous span of nucleotides of a sequence selected as an individual or in any combination from the group consisting of SEQ ID NO: 1-542, the complements thereof, the sequences described in any one or more of Tables 8, 9, 10, 11, 12, 13 and 14 and the complements thereof, wherein said contiguous span is at least 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 75, 100, 200, 500 or 1000 nucleotides in length, to the extent that such a length is consistent with the lengths of the particular Sequence ID. The present invention also relates to polynucleotides hybridizing under stringent or intermediate conditions to a sequence selected from the group consisting of SEQ ID NO: 1-542; and the complements thereof. In addition, the polynucleotides of the invention encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination: Said contiguous span may optionally include the CNS disorder-related biallelic marker in said sequence; Optionally either the original or the alternative allele of Table 9 may be specified as being present at said CNS disorder-related biallelic marker; Optionally either the first or the second allele of Tables 8 or 10 may be specified as being present at said CNS disorder-related biallelic marker; Optionally, said polynucleotide may consists of, or consist essentially of a contiguous span which ranges in length from 8, 10, 12, 15, 18 or 20 to 25, 35, 40, 50, 60, 70, or 80 nucleotides, or be specified as being 12, 15, 18, 20, 25, 35, 40, or 50 nucleotides in length and including a CNS disorder-related biallelic marker of said sequence, and optionally the original allele of Table 9 is present at said biallelic marker; Optionally, said biallelic marker may be within 6, 5, 4, 3, 2, or 1 nucleotides of the center of said polynucleotide or at the center of said polynucleotide; Optionally, the 3' end of said contiguous span may be present at the 3' end of said polynucleotide; Optionally, biallelic marker may be present at the 3' end of said polynucleotide; Optionally, the 3' end of said polynucleotide may be located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500 or 1000 nucleotides upstream of a CNS disorder-related biallelic marker in said sequence, to the extent that such a distance is consistent with the lengths of the particular Sequence ID; Optionally, the 3' end of said polynucleotide may be located 1 nucleotide upstream of a CNS disorder-related biallelic marker in said sequence; and Optionally, said polynucleotide may further comprise a label.

A second embodiment of the invention encompasses any polynucleotide of the invention attached to a solid support. In addition, the polynucleotides of the invention which are attached to a solid support encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said polynucleotides may be specified as attached individually or in groups of at least 2, 5, 8, 10, 12,

15, 20, or 25 distinct polynucleotides of the inventions to a single solid support; Optionally, polynucleotides other than those of the invention may be attached to the same solid support as polynucleotides of the invention; Optionally, when multiple polynucleotides are attached to a solid support they may be attached at random locations, or in an ordered array; Optionally, said ordered array may be addressable.

A third embodiment of the invention encompasses the use of any polynucleotide for, or any polynucleotide for use in, determining the identity of one or more nucleotides at a CNS disorder-related biallelic marker. Microsequencing primers are provided in Table 12. In addition, the polynucleotides of the invention for use in determining the identity of one or more nucleotides at a CNS disorder-related biallelic marker encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination. Optionally, said CNS disorder-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID NO: 1-542; and the complements thereof; Optionally, said polynucleotide may comprise a sequence disclosed in the present specification; Optionally, said polynucleotide may consist of, or consist essentially of any polynucleotide described in the present specification; Optionally, said determining may be performed in a hybridization assay, sequencing assay, microsequencing assay, or an enzyme-based mismatch detection assay; Optionally, said polynucleotide may be attached to a solid support, array, or addressable array; Optionally, said polynucleotide may be labeled.

A fourth embodiment of the invention encompasses the use of any polynucleotide for, or any polynucleotide for use in, amplifying a segment of nucleotides comprising a CNS disorder-related biallelic marker. Amplification primers are provided in Table 13. In addition, the polynucleotides of the invention for use in amplifying a segment of nucleotides comprising a CNS disorder-related biallelic marker encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said CNS disorder-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID 1-130; and the complements thereof; Optionally, said CNS disorder-related biallelic marker may be selected individually or in any combination from the biallelic markers described in Table 7; Optionally, said CNS disorder-related biallelic marker may be selected from the following biallelic markers: 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 99-28106-185, 99-30858-354, 18-20-174, 99-32002-313, 18-31-178, 18-38-395, 99-30853-364, 19-56-140, 19-28-136, 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 16-50-196, 8-19-372, 12-254-180, 10-214-279, 10-217-91, 18-194-130, 18-186-391, 18-198-252, 18-242-300, 20-205-302, 19-58-162, 19-9-45, 19-22-74, 19-88-185, 19-18-310, 19-19-174, 19-17-188, 19-16-127, 99-32148-315, 19-46-322, 99-32131-312, 99-32065-303, 19-44-251, 19-29-303, 18-355-67, 18-353-267, 18-338-305, 16-88-185, 24-243-346, 99-62531-351, 99-54279-152, 99-28171-458, 99-28173-395, 18-186-394, 8-15-126, 99-2409-

298, 99-28722-90 and 99-32306-409; Optionally, said CNS disorder-related biallelic marker may be selected from the following biallelic markers: 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 19-28-136, 16-50-196, 19-58-162, 19-9-45, 20-205-302, 24-243-346, 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 19-17-188 and 19-19-174; Optionally, said polynucleotide may comprise a sequence disclosed in the present specification; Optionally, said polynucleotide may consist of, or consist essentially of any polynucleotide described in the present specification; Optionally, said amplifying may be performed by a PCR or LCR. Optionally, said polynucleotide may be attached to a solid support, array, or addressable array. Optionally, said polynucleotide may be labeled.

A fifth embodiment of the invention encompasses methods of genotyping a biological sample comprising determining the identity of a nucleotide at a CNS disorder-related biallelic marker. In addition, the genotyping methods of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said CNS disorder-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID NO: 1-542, and the complements thereof; Optionally, said CNS disorder-related biallelic marker may be selected individually or in any combination from the biallelic markers described in Table 7; Optionally, said CNS disorder-related biallelic marker may be selected from the following biallelic markers: 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 99-28106-185, 99-30858-354, 18-20-174, 99-32002-313, 18-31-178, 18-38-395, 99-30853-364, 19-56-140, 19-28-136, 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 16-50-196, 8-19-372, 12-254-180, 10-214-279, 10-217-91, 18-194-130, 18-186-391, 18-198-252, 18-242-300, 20-205-302, 19-58-162, 19-9-45, 19-22-74, 19-88-185, 19-18-310, 19-19-174, 19-17-188, 19-16-127, 99-32148-315, 19-46-322, 99-32131-312, 99-32065-303, 19-44-251, 19-29-303, 18-355-67, 18-353-267, 18-338-305, 16-88-185, 24-243-346, 99-62531-351, 99-54279-152, 99-28171-458, 99-28173-395, 18-186-394, 8-15-126, 99-2409-298, 99-28722-90 and 99-32306-409; Optionally, said CNS disorder-related biallelic marker may be selected from the following biallelic markers: 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 19-28-136, 16-50-196, 19-58-162, 19-9-45, 20-205-302, 24-243-346, 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 19-17-188 and 19-19-174; Optionally, said method further comprises determining the identity of a second nucleotide at said biallelic marker, wherein said first nucleotide and second nucleotide are not base paired (by Watson & Crick base pairing) to one another; Optionally, said biological sample is derived from a single individual or subject; Optionally, said method is performed *in vitro*; Optionally, said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome; Optionally, said biological sample is derived from multiple subjects or individuals; Optionally, said method further comprises amplifying a portion of said sequence comprising the biallelic marker prior to said determining step; Optionally, wherein said amplifying is performed

by PCR, LCR, or replication of a recombinant vector comprising an origin of replication and said portion in a host cell; Optionally, wherein said determining is performed by a hybridization assay, sequencing assay, microsequencing assay, or an enzyme-based mismatch detection assay.

A sixth embodiment of the invention comprises methods of estimating the frequency of an allele in a population comprising genotyping individuals from said population for a CNS disorder-related biallelic marker and determining the proportional representation of said biallelic marker in said population. In addition, the methods of estimating the frequency of an allele in a population of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said CNS disorder-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ NO: 1-542; and the complements thereof; Optionally, said CNS disorder-related biallelic marker may be selected from the biallelic markers described in Table 7; Optionally, said CNS disorder-related biallelic marker may be selected from the following biallelic markers: 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 99-28106-185, 99-30858-354, 18-20-174, 99-32002-313, 18-31-178, 18-38-395, 99-30853-364, 19-56-140, 19-28-136, 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 16-50-196, 8-19-372, 12-254-180, 10-214-279, 10-217-91, 18-194-130, 18-186-391, 18-198-252, 18-242-300, 20-205-302, 19-58-162, 19-9-45, 19-22-74, 19-88-185, 19-18-310, 19-19-174, 19-17-188, 19-16-127, 99-32148-315, 19-46-322, 99-32131-312, 99-32065-303, 19-44-251, 19-29-303; 18-355-67, 18-353-267, 18-338-305, 16-88-185, 24-243-346, 99-62531-351, 99-54279-152, 99-28171-458, 99-28173-395, 18-186-394, 8-15-126, 99-2409-298, 99-28722-90 and 99-32306-409; Optionally, said CNS disorder-related biallelic marker may be selected from the following biallelic markers: 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 19-28-136, 16-50-196, 19-58-162, 19-9-45, 20-205-302, 24-243-346, 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 19-17-188 and 19-19-174; Optionally, determining the frequency of a biallelic marker allele in a population may be accomplished by determining the identity of the nucleotides for both copies of said biallelic marker present in the genome of each individual in said population and calculating the proportional representation of said nucleotide at said CNS disorder-related biallelic marker for the population; Optionally, determining the frequency of a biallelic marker allele in a population may be accomplished by performing a genotyping method on a pooled biological sample derived from a representative number of individuals, or each individual, in said population, and calculating the proportional amount of said nucleotide compared with the total.

A seventh embodiment of the invention comprises methods of detecting an association between an allele and a phenotype, comprising the steps of a) determining the frequency of at least one CNS disorder-related biallelic marker allele in a trait positive population, b) determining the frequency of said CNS disorder-related biallelic marker allele in a control

population and; c) determining whether a statistically significant association exists between said genotype and said phenotype. In addition, the methods of detecting an association between an allele and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said CNS disorder-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID NO: 1-542, and the complements thereof; Optionally, said CNS disorder-related biallelic marker may be selected from the biallelic markers described in Table 7; Optionally, said control population may be a trait negative population, or a random population; Optionally, said phenotype is a CNS disorder, a response to an agent acting on a CNS disorder, or side effect to an agent acting on a CNS disorder; Optionally, ~~the~~ identity of the nucleotides at the biallelic markers in everyone of the following sequences: SEQ ID NO: 1-542 is determined in steps a) and b).

An eighth embodiment of the present invention encompasses methods of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising the steps of: a) genotyping each individual in said population for at least one CNS disorder-related biallelic marker, b) genotyping each individual in said population for a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome; and c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency. In addition, the methods of estimating the frequency of a haplotype of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally said haplotype determination method is selected from the group consisting of asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark method, or an expectation maximization algorithm; Optionally, said second biallelic marker is a CNS disorder-related biallelic marker in a sequence selected from the group consisting of the biallelic markers of SEQ ID NO: 1-542, and the complements thereof; Optionally, said CNS disorder-related biallelic markers may be selected individually or in any combination from the biallelic markers described in Table 7; Optionally, said CNS disorder-related biallelic marker may be selected from the following biallelic markers: 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 99-28106-185, 99-30858-354, 18-20-174, 99-32002-313, 18-31-178, 18-38-395, 99-30853-364, 19-56-140, 19-28-136, 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 16-50-196, 8-19-372, 12-254-180, 10-214-279, 10-217-91, 18-194-130, 18-186-391, 18-198-252, 18-242-300, 20-205-302, 19-58-162, 19-9-45, 19-22-74, 19-88-185, 19-18-310, 19-19-174, 19-17-188, 19-16-127, 99-32148-315, 19-46-322, 99-32131-312, 99-32065-303, 19-44-251, 19-29-303, 18-355-67, 18-353-267, 18-338-305, 16-88-185, 24-243-346, 99-62531-351, 99-54279-152, 99-28171-458, 99-28173-395, 18-186-394, 8-15-126, 99-2409-298, 99-28722-90 and 99-32306-4097; Optionally, said CNS disorder-related biallelic marker

may be selected from the following biallelic markers: 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 19-28-136, 16-50-196, 19-58-162, 19-9-45, 20-205-302, 24-243-346, 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 19-17-188 and 19-19-174; Optionally, the identity of the nucleotides at the biallelic markers in everyone of the sequences of SEQ ID NO: 1-542 is determined in steps a) and b).

A ninth embodiment of the present invention encompasses methods of detecting an association between a haplotype and a phenotype, comprising the steps of: a) estimating the frequency of at least one haplotype in a trait positive population according to a method of estimating the frequency of a haplotype of the invention; b) estimating the frequency of said haplotype in a control population according to the method of estimating the frequency of a haplotype of the invention; and c) determining whether a statistically significant association exists between said haplotype and said phenotype. In addition, the methods of detecting an association between a haplotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said CNS disorder-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID NO: 1-542, and the complements thereof; Optionally, said CNS disorder-related biallelic markers may be selected individually or in any combination from the biallelic markers described in Table 7; Optionally, said CNS disorder-related biallelic marker may be selected from the following biallelic markers: 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 99-28106-185, 99-30858-354, 18-20-174, 99-32002-313, 18-31-178, 18-38-395, 99-30853-364, 19-56-140, 19-28-136, 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 16-50-196, 8-19-372, 12-254-180, 10-214-279, 10-217-91, 18-194-130, 18-186-391, 18-198-252, 18-242-300, 20-205-302, 19-58-162, 19-9-45, 19-22-74, 19-88-185, 19-18-310, 19-19-174, 19-17-188, 19-16-127, 99-32148-315, 19-46-322, 99-32131-312, 99-32065-303, 19-44-251, 19-29-303, 18-355-67, 18-353-267, 18-338-305, 16-88-185, 24-243-346, 99-62531-351, 99-54279-152, 99-28171-458, 99-28173-395, 18-186-394, 8-15-126, 99-2409-298, 99-28722-90 and 99-32306-409; Optionally, said CNS disorder-related biallelic marker may be selected from the following biallelic markers: 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 19-28-136, 16-50-196, 19-58-162, 19-9-45, 20-205-302, 24-243-346, 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 19-17-188 and 19-19-174; Optionally, said control population may be a trait negative population, or a random population; Optionally, said phenotype is a CNS disorder, a response to an agent acting on a CNS disorder, or side effect to an agent acting on a CNS disorder; Optionally, the identity of the nucleotides at the biallelic markers in everyone of the following sequences: SEQ ID NO: 1-542 is included in the estimating steps a) and b).

A tenth embodiment of the present invention encompasses polypeptides encoded by SEQ ID NO: 543 or 544, as well as antisense analogs thereof and biologically active and

diagnostically or therapeutically useful fragments and derivatives thereof. The polypeptides of the present invention are of human origin. In accordance with a further aspect of the present invention, there is provided a method for producing such polypeptides by recombinant techniques which comprises culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence encoding a polypeptide of the present invention, under conditions promoting expression of said protein and subsequent recovery of said protein. A further embodiment of the present invention encompasses antibodies against such polypeptides.

An eleventh embodiment of the present invention is a method for using one or more of the polypeptides according to the invention to screen for polypeptide antagonists and/or agonists and/or receptor ligands. A further embodiment of the present invention is a method of using such agonists to activate the polypeptides of the present invention for the treatment of conditions related to the underexpression of the polypeptide of the present invention, preferably depression. In accordance with another aspect of the present invention there is provided a method of using such antagonists for inhibiting the polypeptide of the present invention for treating conditions associated with overexpression of the polypeptides of the present invention.

A twelfth embodiment of the present invention encompasses non-naturally occurring synthetic, isolated and/or recombinant polypeptides which are fragments, consensus fragments and/or sequences having conservative amino acid substitutions, of at least one transmembrane domain, such that the polypeptides of the present invention may bind ligands, or which may also modulate, quantitatively or qualitatively, ligand binding to the polypeptides of the present invention. A further embodiment of the present invention encompasses synthetic or recombinant polypeptides, conservative substitution derivatives thereof, antibodies, anti-idiotype antibodies, compositions and methods that can be useful as potential modulators of CNS-related protein function, by binding to ligands or modulating ligand binding, due to their expected biological properties, which may be used in diagnostic, therapeutic and/or research applications relating to CNS disorders. In yet a further embodiment of the present invention, there is provided synthetic, isolated or recombinant polypeptides which are designed to inhibit or mimic various polypeptides of the invention or fragments thereof, as receptor types and subtypes.

A thirteenth embodiment of the present invention encompasses a diagnostic assay for detecting a disease or susceptibility to a disease related to a mutation in a nucleic acid sequence encoding a polypeptide of the present invention. Preferably said disease is depression.

A fourteenth embodiment of the present invention is a method of administering a drug or a treatment comprising the steps of: a) obtaining a nucleic acid sample from an individual; b) determining the identity of the polymorphic base of at least one CNS disorder-related biallelic marker which is associated with a positive response to the treatment or the drug; or at least one biallelic CNS disorder-related marker which is associated with a negative response to the treatment or the drug; and c) administering the treatment or the drug to the individual if the

nucleic acid sample contains said biallelic marker associated with a positive response to the treatment or the drug or if the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug. In addition, the methods of the present invention for administering a drug or a treatment encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: optionally, said CNS disorder-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ. ID. NO: 1-542 and the complements thereof-, or optionally, the administering step comprises administering the drug or the treatment to the individual if the nucleic acid sample contains said biallelic marker associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

A fifteenth embodiment of the present invention is a method of selecting an individual for inclusion in a clinical trial of a treatment or drug comprising the steps of: a) obtaining a nucleic acid sample from an individual; b) determining the identity of the polymorphic base of at least one CNS disorder-related biallelic marker which is associated with a positive response to the treatment or the drug, or at least one CNS disorder-related biallelic marker which is associated with a negative response to the treatment or the drug in the nucleic acid sample, and c) including the individual in the clinical trial if the nucleic acid sample contains said CNS disorder-related biallelic marker associated with a positive response to the treatment or the drug or if the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug. In addition, the methods of the present invention for selecting an individual for inclusion in a clinical trial of a treatment or drug encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said CNS disorder-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ. ID. NO: 1-542 and the complements thereof, optionally, the including step comprises administering the drug or the treatment to the individual if the nucleic acid sample contains said biallelic marker associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

Additional embodiments are set forth in the Detailed Description of the Invention and in the Examples.

BRIEF DESCRIPTION OF THE TABLES

Tables 7A and 7C are charts containing a list of all of the CNS-related biallelic markers for each gene with an indication of the gene for which the marker is in closest physical proximity, an indication of whether the markers have been validated by microsequencing (with a Y indicating that the markers have been validated by microsequencing and an N indicating that it

has not), and an indication of the identity and frequency of the least common allele determined by genotyping (with a blank left to indicate that the frequency has not yet been reported for some markers).

Tables 7B and 7D contain all of the CNS-related biallelic markers provided in Tables 7A and 7C; however, they are provided in shorter, easier to search sequences of 47 nucleotides. Accordingly, Table 7A begins with SEQ ID NO: 1 and ends with SEQ ID NO: 130, while corresponding Table 7B begins with SEQ ID NO: 131 and ends with SEQ ID NO: 260. Also Table 7C begins with SEQ ID NO: 261 and ends with SEQ ID NO: 401, while corresponding Table 7D begins with SEQ ID NO: 402 and ends with SEQ ID NO: 542. Table 1 contains the first five markers listed in the sequence listing and their corresponding SEQ ID numbers in Tables 7A and 7C to illustrate the relationship between Tables 7A and 7B:

Table 1

BIALLELIC MARKER ID	SEQ ID NO. IN TABLE 7A	BIALLELIC MARKER POSITION IN SEQ ID NO.	SEQ ID NO. IN TABLE 7B	BIALLELIC MARKER POSITION IN SEQ ID NO.
99-27199-207	1	207	131	24
99-27207-117	2	117	132	24
99-27213-53	3	53	133	24
99-27218-333	4	333	134	24
99-28108-233	5	233	135	24

Tables 7B and 7D are the same as Tables 7A and 7C, respectively, in that they are a list of all of the CNS-related biallelic markers for each gene with an indication of the gene for which the marker is in closest physical proximity, an indication of whether the markers have been validated by microsequencing (with a Y indicating that the markers have been validated by microsequencing and an N indicating that it has not), and an indication of the identity and frequency of the least common allele determined by genotyping (with a blank left to indicate that the frequency has not yet been reported for some markers). However, the "Biallelic Marker Position in SEQ ID No." for all of the CNS-related biallelic markers provided in Tables 7B and 7D is position 24 (representing the midpoint of the 47mers that make up Tables 7B and 7D).

Tables 8, 9, and 10 are charts containing lists of the CNS disorder-related biallelic markers. Each marker is described by indicating its SEQ ID, the biallelic marker ID, and the two most common alleles. Table 8 is a chart containing a list of biallelic markers surrounded by preferred sequences. In the column labeled, "POSITION RANGE OF PREFERRED SEQUENCE" of Table 8 regions of particularly preferred sequences are listed for each SEQ ID, which contain a CNS disorder-related biallelic marker, as well as particularly preferred regions

of sequences that do not contain a CNS disorder-related biallelic marker but, which are in sufficiently close proximity to a CNS disorder-related biallelic marker to be useful as amplification or sequencing primers.

Table 11 is a chart listing particular sequences that are useful for designing some of the primers and probes of the invention. Each sequence is described by indicating its Sequence ID and the positions of the first and last nucleotides (position range) of the particular sequence in the Sequence ID.

Table 12 is a chart listing microsequencing primers which have been used to genotype CNS disorder-related biallelic markers (indicated by an *) and other preferred microsequencing primers for use in genotyping CNS disorder-related biallelic markers. Each of the primers which falls within the strand of nucleotides included in the Sequence Listing are described by indicating their Sequence ID number and the positions of the first and last nucleotides (position range) of the primers in the Sequence ID. Since the sequences in the Sequence Listing are single stranded and half the possible microsequencing primers are composed of nucleotide sequences from the complementary strand, the primers that are composed of nucleotides in the complementary strand are described by indicating their SEQ ID numbers and the positions of the first and last nucleotides to which they are complementary (complementary position range) in the Sequence ID.

Table 13 is a chart listing amplification primers which have been used to amplify polynucleotides containing one or more CNS disorder-related biallelic markers. Each of the primers which falls within the strand of nucleotides included in the Sequence Listing are described by indicating their Sequence ID number and the positions of the first and last nucleotides (position range) of the primers in the Sequence ID. Since the sequences in the Sequence Listing are single stranded and half the possible amplification primers are composed of nucleotide sequences from the complementary strand, the primers that are composed of nucleotides in the complementary strand are defined by the SEQ ID numbers and the positions of the first and last nucleotides to which they are complementary (complementary position range) in the Sequence ID.

Table 14 is a chart listing preferred probes useful in genotyping CNS disorder-related biallelic markers by hybridization assays. The probes are 25-mers with a CNS disorder-related biallelic markers in the center position, and described by indicating their Sequence ID number and the positions of the first and last nucleotides (position range) of the probes in the Sequence ID. The probes complementary to the sequences in each position range in each Sequence ID are also understood to be a part of this preferred list even though they are not specified separately.

Table 15 is a table showing the results of single marker association tests between both biallelic marker alleles and genotypes of candidate genes and major depression.

Table 16 is a table showing the results of the LR rank of haplotypes using combinations of 2, 3 and 4 biallelic markers from each gene.

Table 17 is a table showing the rank of permutation tests for individual haplotypes confirming the statistical significance of the association between biallelic marker haplotypes from the candidate genes and major depression.

Table 18 is a table showing the results of single marker association tests between both biallelic marker alleles and genotypes of candidate genes and major depression using additional markers and a new population set as described in Example 4.

Table 19 is a table showing the results of the LR rank of haplotypes using combinations of 2, 3 and 4 biallelic markers from additional candidate genes and using data from a new population set as described in Example 4.

Table 20 is a table showing the rank of permutation tests for individual haplotypes from Table 19 confirming the statistical significance of the association between biallelic marker haplotypes from additional candidate genes and major depression.

DETAILED DESCRIPTION OF THE INVENTION

I. Candidate Genes of the Present Invention

Different approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. Genome-wide association studies rely on the screening of genetic markers evenly spaced and covering the entire genome. Candidate region association studies rely on the screening of genetic markers evenly spaced covering a region identified as linked to the trait of interest. The candidate gene approach is based on the study of genetic markers specifically derived from genes potentially involved in the pathophysiology of a disease. In the present invention, genes involved in the central nervous system and/or the endocrine system have been chosen as candidate genes. The candidate genes of the present invention are listed in Table 2.

Table 2

Candidate Gene Name	Gene Symbol	Description
Serotonin receptor 6	5HTR6	A postsynaptic serotonin receptor.
Serotonin receptor 7	5HTR7	A postsynaptic serotonin receptor.
Neuronal nicotinic acid receptor α7	CHRNA7	An ion channel in the reward pathway.
Corticotrophin releasing factor receptor 1	CRFR1	A corticotrophin releasing factor receptor in the hypothalamus-pituitary-adrenal axis.
Mineralocorticoid receptor	MLR	A mineralocorticoid receptor.

Corticotrophin releasing factor receptor 2	CRFR2	A corticotrophin releasing factor receptor in the hypothalamus-pituitary-adrenal axis.
Glucocorticoid receptor	GRL	A glucocorticoid receptor in the hypothalamus-pituitary-adrenal axis.
Monoamine oxidase A	MAOA	Key enzyme in catecholamine metabolism.
Monoamine oxidase B	MAOB	Key enzyme in catecholamine metabolism.
Serotonin receptor 2C	5HTR2c	Postsynaptic receptor for serotonin.
Tyrosine hydroxylase	TH	The rate-limiting enzyme in the synthesis of dopamine and norepinephrine.
Corticotrophin releasing factor	CRF	A hormone released by the hypothalamus that stimulates the release of corticotrophin by the anterior pituitary gland.
Dopamine receptor 4	DRD4	A postsynaptic dopamine receptor.
Serotonin transporter	5HTT	A presynaptic membrane receptor that serves as a reuptake mechanism for serotonin.
Dopamine receptor 3	DRD3	A postsynaptic dopamine receptor.
Cytochrome P450 3A4	CYP3A4	A principal drug metabolizing enzyme.
Norepinephrine transporter	NET	A membrane protein responsible for termination of the action of synaptic norepinephrine.
Neurokinin or tachykinin receptor 1	NK1/TACR1	A receptor for the neuropeptide substance P.
Neuropeptide Y1 receptor	NPY1R	A receptor for the neuropeptide Y1. Belongs to family of g-protein coupled receptors with it highest similarity to tachykinins receptors.
Dopamine receptor 2	DRD2	G protein-coupled dopamine receptor.
Guanine nucleotide binding protein, β3	Gbeta3	An important component of cAMP mediated signaling pathways.
Wolfram Syndrome 1 gene	WFS1	A gene that plays a role in the etiology of Wolfram syndrome.
Beta 1 adrenergic receptor	ADRB1R	An important component of the norepinephrine signaling pathway. Antidepressants are known to suppress expression.
Brain derived neurotrophic factor	BDNF	A protein known to affect the differentiation of dopaminergic and serotonergic neurons.

		Increased by antidepressants and electro-convulsive therapy.
Orphan G-protein coupled receptor	HM74	A putative chemokine receptor.
Vasopressin receptor 1A	AVPR1A	A receptor that stimulates adrenocorticotrophic hormone (ACTH) release in the anterior pituitary.
Serotonin receptor 1-A	5HT1A	A receptor that is misregulated in depression, as well as anxiety and stress.
Growth associated protein 43	GAP43	A protein known to play a role in synaptic plasticity. Increased levels in suicide victims.
Guanine nucleotide binding protein, α subunit, olfactory type	GOLF (GNAL)	A protein known to play a role in signaling: possibly in cAMP mediated signaling pathways and norepinephrine-related pathways.
Clock protein	CLOCK	A protein associated with sleeping patterns in humans.
Corticotrophin hormone binding protein	CRHBP	A protein capable of binding to corticotrophin releasing factor.
Dopamine transporter	DAT (SLC6A3)	A protein involved in the re-uptake of dopamine.
Phosphodiesterase type 4b	PDE4b	An enzyme believed to be involved in mediating central nervous system effects of therapeutic agents ranging from antidepressants to anti-inflammatory agents.
Catechol O-methyl transferase	COMT	An enzyme that catalyzes the transfer of a methyl group from s-adenosylmethionine to a catecholamine such as dopamine, epinephrine, or norepinephrine.
Melanin concentrating hormone receptor	SLC1	A transmembrane protein that serves as the functional receptor of melanin concentrating hormone.
Transcription factor	SEF2-1B (TCF4)	A transcription factor that binds to the e-box present in the somatostatin receptor 2 initiator element (sstr2-inr) to activate transcription (by similarity).

Heat shock protein	HSP70	A protein believed to interact with polypeptides during a variety of assembly processes in such a way as to prevent the formation of nonfunctional structures.
GABA-A receptor subunit	GABRG2	A receptor known to mediate inhibitory neurotransmission, complexing with DRD5 and promoting mutually inhibitory functional interactions between these receptor systems, putatively involved in the physiological dependence on alcohol, and in the maintenance of psychomotor disease states.
GABA-A receptor subunit 5	GABRA5	A receptor known to be part of the ligand-gated ionic channels protein family. Associated with bipolar disorder.

Both the central nervous system and the endocrine system play an important role in the pathophysiology of CNS disorders, moreover, these systems contain important drug targets and genetic polymorphisms in these genes are highly relevant in the response to a number of drugs. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular disease when some information concerning the pathophysiology of the disorder is available as is the case for many CNS disorders. However, it should be noted that all of the biallelic markers disclosed in the instant application can be employed as part of genome-wide association studies or as part of candidate region association studies and such uses are specifically contemplated in the present invention and claims. All of the markers are known to be in close proximity to the genes with which they are listed in Table 7. For a portion of the markers, the precise position of the marker with respect to the various coding and non-coding elements of the genes has also been determined.

II. Definitions

Before describing the invention in greater detail, the following definitions are set forth to illustrate and define the meaning and scope of the terms used to describe the invention herein.

As used interchangeably herein, the terms "nucleic acid molecule", "oligonucleotide", and "polynucleotide", unless specifically stated otherwise, include RNA or, DNA (either single or double stranded, coding, complementary or antisense), or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form (although each of the above species may be particularly specified). The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in

single-stranded or duplex form. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. Although the term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. Preferred modifications of the present invention include, but are not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v) ybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, and 2,6-diaminopurine. The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art. Methylenemethylimino linked oligonucleosides as well as mixed backbone compounds having, may be prepared as described in U.S. Pat. Nos. 5,378,825; 5,386,023; 5,489,677; 5,602,240; and 5,610,289. Formacetal and thioformacetal linked oligonucleosides may be prepared as described in U.S. Pat. Nos. 5,264,562 and 5,264,564. Ethylene oxide linked oligonucleosides may be prepared as described in U.S. Pat. No. 5,223,618. Phosphinate oligonucleotides may be prepared as described in U.S. Pat. No. 5,508,270.. Alkyl phosphonate oligonucleotides may be prepared as described in U.S. Pat. No. 4,469,863. 3'-Deoxy-3'-methylene phosphonate oligonucleotides may be prepared as described in U.S. Pat. Nos. 5,610,289 or 5,625,050. Phosphoramidite oligonucleotides may be prepared as described in U.S. Pat. No. 5,256,775 or U.S. Pat. No. 5,366,878. Alkylphosphonothioate oligonucleotides may be prepared as described in published PCT applications WO 94/17093 and WO 94/02499. 3'-Deoxy-3'-amino phosphoramidate oligonucleotides may be prepared as described in U.S. Pat. No. 5,476,925. Phosphotriester

oligonucleotides may be prepared as described in U.S. Pat. No. 5,023,243. Borano phosphate oligonucleotides may be prepared as described in U.S. Pat. Nos. 5,130,302 and 5,177,198. The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art.

The term "isolated" further requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated. Specifically excluded from the definition of "isolated" are: naturally-occurring chromosomes (such as chromosome spreads), artificial chromosome libraries, genomic libraries, and cDNA libraries that exist either as an *in vitro* nucleic acid molecule preparation or as a transfected/transformed host cell preparation, wherein the host cells are either an *in vitro* heterogeneous preparation or plated as a heterogeneous population of single colonies. Also specifically excluded are the above libraries wherein a specified polynucleotide of the present invention makes up less than 5% of the number of nucleic acid molecule inserts in the vector molecules. Further specifically excluded are whole cell genomic DNA or whole cell RNA or mRNA preparations (including said whole cell preparations which are mechanically sheared or enzymatically digested). Further specifically excluded are the above whole cell preparations as either an *in vitro* preparation or as a heterogeneous mixture separated by electrophoresis (including blot transfers of the same) wherein the polynucleotide of the invention has not further been separated from the heterologous polynucleotides in the electrophoresis medium (e.g., further separating by excising a single band from a heterogeneous band population in an agarose gel or nylon blot).

As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual 5' EST clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately 10^4-10^6 fold purification of the native message. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. Alternatively, purification may be expressed as "at least" a percent purity relative to heterologous polynucleotides (DNA, RNA

or both). As a preferred embodiment, the polynucleotides of the present invention are at least; 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 96%, 98%, 99%, or 100% pure relative to heterologous polynucleotides. As a further preferred embodiment the polynucleotides have an "at least" purity ranging from any number, to the thousandth position, between 90% and 100% (e.g., 5' EST at least 99.995% pure) relative to heterologous polynucleotides. Additionally, purity of the polynucleotides may be expressed as a percentage (as described above) relative to all materials and compounds other than the carrier solution. Each number, to the thousandth position, may be claimed as individual species of purity.

The term "primer" denotes a specific oligonucleotide sequence which is complementary to a target nucleotide sequence and used to hybridize to the target nucleotide sequence. A primer serves as an initiation point for nucleotide polymerization catalyzed by DNA polymerase, RNA polymerase or reverse transcriptase.

The term "probe" denotes a defined nucleic acid segment (or nucleotide analog segment, e.g., polynucleotide as defined herein) which can be used to identify a specific polynucleotide sequence present in samples, said nucleic acid segment comprising a nucleotide sequence complementary of the specific polynucleotide sequence to be identified.

The term "CNS disorder" refers to any condition linked to dysfunction of the central nervous system which is known in the art. A CNS disorder includes dysfunction of one or several physiological systems contributing to the function of the central nervous system, which includes the endocrine system and the peripheral nervous system. A CNS disorder further refers to disorders in neurotransmitter synthesis and degradation, neurotransmitter function, neurotransmitter receptor function, neurotransmitter signal transduction, neurotransmitter transporter function, motor neuron function, hormone synthesis and degradation, hormone function, hormone receptor function and hormone signal transduction. "CNS disorders" include mood disorders such as depression, bipolar disorder, anxiety, attention deficit disorder and schizophrenia. "CNS disorders" also include neurodegenerative disorders such as Parkinson's disease, Huntington's disease, Pick's disease, progressive supranuclear palsy, Lewy body dementia and Wolfram syndrome (diabetes insipidus, diabetes mellitus, optic atrophy and deafness). "CNS disorders" also include disorders of movement such as motor neuron disease as well as diseases involving the intellect such as Alzheimer's disease, Wernicke's encephalopathy and Jakob-Creutzfeldt disease. "CNS disorders" further include other disorders such as coma, head injury, cerebral infarction, epilepsy, alcoholism and states of mental retardation. All of the possible CNS disorders listed herein are included in, or may be excluded from, the present invention as individual species.

The term "depression" as used herein refers to both unipolar major depression (or major depressive disorder) and bipolar disorder.

An "agent acting on a CNS disorder" includes any drug or compound known in the art that addresses, reduces or alleviates one or more symptoms of a CNS disorder. "Agents acting on a CNS disorder" includes any drug or a compound modulating the activity or concentration of an enzyme or regulatory molecule involved in a CNS disorder that is known in the art. An agent acting on a CNS disorder includes but is not limited to tyrosine hydroxylase, monoamine oxidase A/B, dopamine β-hydroxylase, aldehyde dehydrogenase, phenylethanolamine N-methyltransferase, catechol o-methyltransferase, tryptophan hydroxylase, acetyl coenzyme A, proteinases, oestrogens, glucocorticoids, mineralocorticoids, nicotine, substance P and precursors to neurotransmitters such as tryptophan. "Agent acting on a CNS disorder" further refers to compounds modulating the synthesis, degradation, reuptake and action of neurotransmitters and hormones such as tricyclic antidepressants (TCAs), monoamine oxidase inhibitors (MAOIs), serotonin selective reuptake inhibitors (SSRIs), selective norepinephrine reuptake inhibitor (NRI) such as reboxetine, dual serotonin and norepinephrine reuptake inhibitor (SNRI), serotonin-2 antagonist/reuptake inhibitors (SARIs), noradrenergic and specific serotonergic antidepressants (NaSSAs), drugs that cause increased dopamine release in the brain such as levodopa, dopamine receptor agonists such as bromocriptine, dopamine antagonists such as metoclopramide, neuroleptic drugs such as phenothiazines, adrenergic agonists such as clonidine, N-methyl-D-asparate antagonists such as phencyclidine, anticholinergic compounds, benzodiazepine drugs and anxiolytic compounds. Preferably, "agent acting on a CNS disorder" refers to the antidepressant drug Reboxetine.

The terms "response to an agent acting on a CNS disorder" refer to drug efficacy, including but not limited to the ability to metabolize a compound, the ability to convert a pro-drug to an active drug, and to the pharmacokinetics (absorption, distribution, elimination) and the pharmacodynamics (receptor-related) of a drug in an individual.

The terms "side effects to an agent acting on a CNS disorder" refer to adverse effects of therapy resulting from extensions of the principal pharmacological action of the drug or to idiosyncratic adverse reactions resulting from an interaction of the drug with unique host factors. "Side effects to an agent acting on a CNS disorder" include, but are not limited to autonomic side effects such as orthostatic hypotension, blurred vision, dry mouth, nasal congestion and constipation. "Side effects to an agent acting on a CNS disorder" also include anxiety, sleep disturbances, sexual dysfunction, gastrointestinal disturbances, nausea, diarrhea, orthostasis, dizziness, sedation, hypertension and shock.

The terms "trait" and "phenotype" are used interchangeably herein and refer to any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to a disease for example. Typically the terms "trait" or "phenotype" are used herein to refer to symptoms of, or susceptibility to a CNS disorder; or to refer to an individual's

response to an agent acting on a CNS disorder; or to refer to symptoms of, or susceptibility to side effects to an agent acting on a CNS disorder.

The term "allele" is used herein to refer to variants of a nucleotide sequence. A biallelic polymorphism has two forms. Typically the first identified allele is designated as the original allele whereas other alleles are designated as alternative alleles. Diploid organisms may be homozygous or heterozygous for an allelic form.

The term "heterozygosity rate" is used herein to refer to the incidence of individuals in a population, which are heterozygous at a particular allele. In a biallelic system the heterozygosity rate is on average equal to $2P_a(1-P_a)$, where P_a is the frequency of the least common allele. In order to be useful in genetic studies a genetic marker should have an adequate level of heterozygosity to allow a reasonable probability that a randomly selected person will be heterozygous.

The term "genotype" as used herein refers to the identity of the alleles present in an individual or a sample. In the context of the present invention a genotype preferably refers to the description of the biallelic marker alleles present in an individual or a sample. The term "genotyping" a sample or an individual for a biallelic marker consists of determining the specific allele or the specific nucleotide carried by an individual at a biallelic marker.

The term "mutation" as used herein refers to a difference in DNA sequence between or among different genomes or individuals which has a frequency below 1%.

The term "haplotype" refers to a combination of alleles present in an individual or a sample. In the context of the present invention a haplotype preferably refers to a combination of biallelic marker alleles found in a given individual and which may be associated with a phenotype.

The term "polymorphism" as used herein refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals. "Polymorphic" refers to the condition in which two or more variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation occurs. A single nucleotide polymorphism is a single base pair change. Typically a single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide, also give rise to single nucleotide polymorphisms. In the context of the present invention "single nucleotide polymorphism" preferably refers to a single nucleotide substitution. Typically, between different genomes or between different individuals, the polymorphic site may be occupied by two different nucleotides.

The terms "biallelic polymorphism" and "biallelic marker" are used interchangeably herein to refer to a polymorphism having two alleles at a fairly high frequency in the population, preferably a single nucleotide polymorphism. A "biallelic marker allele" refers to the nucleotide

variants present at a biallelic marker site. Typically the frequency of the less common allele of the biallelic markers of the present invention has been validated to be greater than 1%, preferably the frequency is greater than 10%, more preferably the frequency is at least 20% (i.e. heterozygosity rate of at least 0.32), even more preferably the frequency is at least 30% (i.e. heterozygosity rate of at least 0.42). A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker."

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotide positions in the middle of the polynucleotide would be considered to be within 2 nucleotides of the center, and so on. When a polynucleotide has an even number of nucleotides, there would be a bond and not a nucleotide at the center of the polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1 nucleotide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on. For polymorphisms which involve the substitution, insertion or deletion of 1 or more nucleotides, the polymorphism, allele or biallelic marker is "at the center" of a polynucleotide if the difference between the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 3' end of the polynucleotide, and the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 5' end of the polynucleotide is zero or one nucleotide. If this difference is 0 to 3, then the polymorphism is considered to be "within 1 nucleotide of the center." If the difference is 0 to 5, the polymorphism is considered to be "within 2 nucleotides of the center." If the difference is 0 to 7, the polymorphism is considered to be "within 3 nucleotides of the center," and so on. For polymorphisms which involve the substitution, insertion or deletion of 1 or more nucleotides, the polymorphism, allele or biallelic marker is "at the center" of a polynucleotide if the difference between the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 3' end of the polynucleotide, and the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 5' end of the polynucleotide is zero or one nucleotide. If this difference is 0 to 3, then the polymorphism is considered to be "within 1 nucleotide of the center." If the difference is 0 to 5, the polymorphism is considered to be "within 2 nucleotides of the center." If the difference is 0 to 7, the polymorphism is considered to be "within 3 nucleotides of the center," and so on.

The term "upstream" is used herein to refer to a location which is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another be virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, L., *Biochemistry,* 4th edition, 1995).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. This term is applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

As used herein the term "CNS disorder-related biallelic marker" relates to a set of biallelic markers in linkage disequilibrium with genes disclosed in Tables 7(A-D) which express proteins that are involved in the pathophysiology CNS disorders. The term CNS disorder-related biallelic marker encompasses all of the biallelic markers disclosed in Tables 7(A-D). The preferred CNS disorder-related biallelic marker alleles of the present invention include each one of the alleles described in Tables 7, 8, 9, and 10 individually or in groups consisting of all the possible combinations of the alleles included in Tables 7, 8, 9, and 10. In addition, Table 7 may include Tables 7A-7D, or Tables 7A, 7B, 7C or 7D as individual embodiments of the present invention or in any combination of the four.

The term "sequence described in Table 8" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Table 8. The SEQ ID that contains each "sequence described in Table 8" is provided in the column labeled, "SEQ ID NO." The range of nucleotide positions within the Sequence ID of which each sequence consists is provided in the same row as the Sequence ID in a column labeled, "POSITION RANGE OF PREFERRED SEQUENCE". It should be noted that some of the Sequence ID numbers have multiple sequence ranges listed, because they contain multiple "sequences described in Table 8." Unless otherwise noted the term "sequence described in Table 8" is to be construed as encompassing sequences that contain either of the two alleles listed in the columns labeled, "1ST ALLELE" and "2ND ALLELE" at the position identified in field <222> of the allele feature in the appended Sequence Listing for each Sequence ID number referenced in Table 8.

The term "sequence described in Table 9" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Table 9. Unless otherwise noted, the "sequences described in Table 9" consist of the entire sequence of each Sequence ID provided in the column labeled, "SEQ ID NO." Also unless otherwise noted the term "sequence described in Table 9" is to be construed as encompassing sequences that contain either of the two

alleles listed in the columns labeled, "ORIGINAL ALLELE" and "ALTERNATIVE ALLELE" at the position identified in field <222> of the allele feature in the appended Sequence Listing for each Sequence ID number referenced in Table 9.

The term "sequence described in Table 10" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Table 10. Unless otherwise noted, the "sequences described in Table 10" consist of the entire sequence of each Sequence ID provided in the column labeled, "SEQ ID NO." Also unless otherwise noted the term "sequence described in Table 10" is to be construed as encompassing sequences that contain either of the two alleles listed in the columns labeled, "1ST ALLELE" and "2ND ALLELE" at the position identified in field <222> of the allele feature in the appended Sequence Listing for each Sequence ID number referenced in Table 10.

The term "sequence described in Table 11" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Table 11. The SEQ ID that contains each "sequence described in Table 11" is provided in the column labeled, "SEQ ID NO." The range of nucleotide positions within the Sequence ID of which each sequence consists is provided in the same row as the Sequence ID in a column labeled, "POSITION RANGE OF PREFERRED SEQUENCE". It should be noted that some of the Sequence ID numbers have multiple sequence ranges listed, because they contain multiple "sequences described in Table 11."

The term "sequence described in Table 12" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Table 12. The SEQ ID that contains each "sequence described in Table 12" is provided in the column labeled, "SEQ ID NO." The range of nucleotide positions within the Sequence ID of which half of the sequences consists is provided in the same row as the Sequence ID in a column labeled, "POSITION RANGE OF MICROSEQUENCING PRIMERS". The remaining half of the sequences described in Table 12 are complementary to the range of nucleotide positions within the Sequence ID provided in the same row as the Sequence ID in a column labeled, "COMPLEMENTARY POSITION RANGE OF MICROSEQUENCING PRIMERS".

The term "sequence described in Table 13" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Table 13. The SEQ ID that contains each "sequence described in Table 13" is provided in the column labeled, "SEQ ID NO." The range of nucleotide positions within the Sequence ID of which half of the sequences consists is provided in the same row as the Sequence ID in a column labeled, "POSITION RANGE OF AMPLIFICATION PRIMERS". The remaining half of the sequences described in Table 13 are complementary to the range of nucleotide positions within the Sequence ID provided in the same row as the Sequence ID in a column labeled, "COMPLEMENTARY POSITION RANGE OF AMPLIFICATION PRIMERS".

The term "sequence described in Table 14" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Table 14. The SEQ ID that contains each "sequence described in Table 14" is provided in the column labeled, "SEQ ID NO.". The range of nucleotide positions within the Sequence ID of which each sequence consists is provided in the same row as the Sequence ID in a column labeled, "POSITION RANGE OF PROBES". The sequences which are complementary to the ranges listed in the column labeled, "POSITION RANGE OF PROBES" are also encompassed by the term, "sequence described in Table 14." Unless otherwise noted the term "sequence described in Table 14" is to be construed as encompassing sequences that contain either of the two alleles listed in the allele feature in the appended Sequence Listing for each Sequence ID number referenced in Table 14.

The terms "biallelic marker described in Table" and "allele described in Table" are used herein to refer to any or all alleles which are listed in the allele feature in the appended Sequence Listing for each Sequence ID number referenced in the particular Table being mentioned.

The following abbreviations are used in this disclosure: serotonin receptor 6 gene is abbreviated 5HTR6; serotonin receptor 7 gene is abbreviated 5HTR7; neuronal nicotinic acid receptor α7 gene is abbreviated CHRNA7; corticotrophin releasing factor receptor 1gene is abbreviated CRFR1; mineralocorticoid receptor gene is abbreviated MLR; corticotrophin releasing factor receptor 2 gene is abbreviated CRFR2; glucocorticoid receptor gene is abbreviated GRL; monoamine oxidases A and B genes are abbreviated MAOA/B; serotonin receptor 2C gene is abbreviated 5HTR2c; tyrosine hydroxylase gene is abbreviated TH; corticotrophin releasing factor gene is abbreviated CRF; dopamine receptor 4 gene is abbreviated DRD4; serotonin transporter gene is abbreviated 5HTT; dopamine receptor 3 gene is abbreviated DRD3; cytochrome P450 3A4 gene is abbreviated CYP3A4; norepinephrine transporter gene is abbreviated NET; neurokinin or tachykinin receptor 1 gene is abbreviated NK1/TACR1; dopamine receptor 4 gene is abbreviated DRD2; guanine nucleotide binding protein, β3 gene is abbreviated Gbeta3; Wolfram Syndrome 1 gene is abbreviated WFS1; Beta 1 adrenergic receptor gene is abbreviated ADRB1R; Brain derived neurotrophic factor gene is abbreviated BDNF; Orphan G-protein coupled receptor gene is abbreviated HM74; Vasopressin receptor 1A gene is abbreviated AVPR1A; Serotonin receptor 1-A gene is abbreviated 5HT1A; Growth associated protien 43 gene is abbreviated GAP43; Guanine nucleotide binding protein, α subunit, olfactory type gene is abbreviated GOLF (GNAL); Clock protein gene is abbreviated CLOCK; Corticotrophin hormone binding protein gene is abbreviated CRHBP; Dopamine transporter gene is abbreviated DAT (SLC6A3); Phosphodiesterase type 4b gene is abbreviated PDE4b; Catechol O-methyl transferase gene is abbreviated COMT; Melanin concentrating hormone receptor gene is abbreviated SLC1; Transcription factor gene is abbreviated SEF2-1B (TCF4); Heat shock protein gene is abbreviated HSP70; GABA-A receptor subunit gene is abbreviated GABRG2; and GABA-A receptor subunit 5 gene is abbreviated GABRA5.

III. Biallelic Markers and Polynucleotides Comprising Biallelic Markers

A. Advantages of the Biallelic Markers of the Present Invention

The CNS disorder-related biallelic markers of the present invention offer a number of important advantages over other genetic markers such as RFLP (Restriction fragment length polymorphism) and VNTR (Variable Number of Tandem Repeats) markers.

The first generation of markers, were RFLPs, which are variations that modify the length of a restriction fragment. But methods used to identify and to type RFLPs are relatively wasteful of materials, effort, and time. The second generation of genetic markers were VNTRs, which can be categorized as either minisatellites or microsatellites. Minisatellites are tandemly repeated DNA sequences present in units of 5-50 repeats which are distributed along regions of the human chromosomes ranging from 0.1 to 20 kilobases in length. Since they present many possible alleles, their informative content is very high. Minisatellites are scored by performing Southern blots to identify the number of tandem repeats present in a nucleic acid sample from the individual being tested. However, there are only 10^4 potential VNTRs that can be typed by Southern blotting. Moreover, both RFLP and VNTR markers are costly and time-consuming to develop and assay in large numbers.

Single nucleotide polymorphism or biallelic markers can be used in the same manner as RFLPs and VNTRs but offer several advantages. Single nucleotide polymorphisms are densely spaced in the human genome and represent the most frequent type of variation. An estimated number of more than 10^7 sites are scattered along the $3x10^9$ base pairs of the human genome. Therefore, single nucleotide polymorphism occur at a greater frequency and with greater uniformity than RFLP or VNTR markers which means that there is a greater probability that such a marker will be found in close proximity to a genetic locus of interest. Single nucleotide polymorphisms are less variable than VNTR markers but are mutationally more stable.

Also, the different forms of a characterized single nucleotide polymorphism, such as the biallelic markers of the present invention, are often easier to distinguish and can therefore be typed easily on a routine basis. Biallelic markers have single nucleotide based alleles and they have only two common alleles, which allows highly parallel detection and automated scoring. The biallelic markers of the present invention offer the possibility of rapid, high-throughput genotyping of a large number of individuals.

Biallelic markers are densely spaced in the genome, sufficiently informative and can be assayed in large numbers. The combined effects of these advantages make biallelic markers extremely valuable in genetic studies. Biallelic markers can be used in linkage studies in families, in allele sharing methods, in linkage disequilibrium studies in populations, in association studies of case-control populations. An important aspect of the present invention is that biallelic markers allow association studies to be performed to identify genes involved in

complex traits. Association studies examine the frequency of marker alleles in unrelated case- and control-populations and are generally employed in the detection of polygenic or sporadic traits. Association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families (linkage studies). Biallelic markers in different genes can be screened in parallel for direct association with disease or response to a treatment. This multiple gene approach is a powerful tool for a variety of human genetic studies as it provides the necessary statistical power to examine the synergistic effect of multiple genetic factors on a particular phenotype, drug response, sporadic trait, or disease state with a complex genetic etiology.

B. Polynucleotides of the Present Invention

The present invention encompasses polynucleotides for use as primers and probes in the methods of the invention. These polynucleotides may consist of, consist essentially of, or comprise a contiguous span of nucleotides of a sequence from any sequence in the Sequence Listing as well as sequences which are complementary thereto ("complements thereof"). The "contiguous span" may be at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000, 2000 or 3000 nucleotides in length, to the extent that a contiguous span of these lengths is consistent with the lengths of the particular Sequence ID. It should be noted that the polynucleotides of the present invention are not limited to having the exact flanking sequences surrounding the polymorphic bases which, are enumerated in the Sequence Listing. Rather, it will be appreciated that the flanking sequences surrounding the biallelic markers, or any of the primers of probes of the invention which, are more distant from the markers, may be lengthened or shortened to any extent compatible with their intended use and the present invention specifically contemplates such sequences. It will be appreciated that the polynucleotides referred to in the Sequence Listing may be of any length compatible with their intended use. Also the flanking regions outside of the contiguous span need not be homologous to native flanking sequences which actually occur in human subjects. The addition of any nucleotide sequence, which is compatible with the nucleotides intended use is specifically contemplated. The contiguous span may optionally include the CNS disorder-related biallelic marker in said sequence. Biallelic markers generally consist of a polymorphism at one single base position. Each biallelic marker therefore corresponds to two forms of a polynucleotide sequence which, when compared with one another, present a nucleotide modification at one position. Usually, the nucleotide modification involves the substitution of one nucleotide for another. Optionally either the original or the alternative allele of the biallelic markers disclosed in Table 9, or the first or second allele disclosed in Table 8 and 10 may be specified as being present at the CNS disorder-related biallelic marker.

The invention also relates to polynucleotides that hybridize, under conditions of high or intermediate stringency, to a polynucleotide of a sequence from any sequence in the Sequence

Listing as well as sequences, which are complementary thereto. Preferably such polynucleotides are at least 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000, 2000 or 3000 nucleotides in length, to the extent that a polynucleotide of these lengths is consistent with the lengths of the particular Sequence ID. Preferred polynucleotides comprise a CNS disorder-related biallelic marker. Optionally either the original or the alternative allele of the biallelic markers disclosed in Table 9 may be specified as being present at the CNS disorder-related biallelic marker. Conditions of high and intermediate stringency are further described herein.

The preferred polynucleotides of the invention include the sequence ranges included in any one the sequence ranges of Tables 8 and 11 to 14 individually or in groups consisting of all the possible combinations ~~of~~ the ranges of included in Tables 8, and 11 to 14. The preferred polynucleotides of the invention also include fragments of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 consecutive nucleotides of the sequence ranges included in any one of the sequence ranges of Tables 9, and 12 to 15 to the extent that fragments of these lengths are consistent with the lengths of the particular sequence range. The preferred polynucleotides of the invention also include fragments of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 consecutive nucleotides of the sequence complementary to the sequence ranges included in any one of the sequence ranges of Tables 8 and 11 to 14 to the extent that fragments of these lengths are consistent with the lengths of the particular sequence range.

The primers of the present invention may be designed from the disclosed sequences for any method known in the art. A preferred set of primers is fashioned such that the 3' end of the contiguous span of identity with the sequences of the Sequence Listing is present at the 3' end of the primer. Such a configuration allows the 3' end of the primer to hybridize to a selected nucleic acid sequence and dramatically increases the efficiency of the primer for amplification or sequencing reactions. In a preferred set of primers the contiguous span is found in one of the sequences described in Table 11. Allele specific primers may be designed such that a biallelic marker is at the 3' end of the contiguous span and the contiguous span is present at the 3' end of the primer. Such allele specific primers tend to selectively prime an amplification or sequencing reaction so long as they are used with a nucleic acid sample that contains one of the two alleles present at a biallelic marker. The 3' end of primer of the invention may be located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, 1000, 2000 or 3000 to the extent that this distance is consistent with the particular Sequence ID, nucleotides upstream of a CNS disorder-related biallelic marker in said sequence or at any other location which is appropriate for their intended use in sequencing, amplification or the location of novel sequences or markers. A list of preferred amplification primers is disclosed in Table 13. Primers with their 3' ends located 1 nucleotide upstream of a CNS disorder-related biallelic marker have a special utility as microsequencing assays. Preferred microsequencing primers are described in Tables 12.

The probes of the present invention may be designed from the disclosed sequences for any method known in the art, particularly methods which allow for testing if a particular sequence or marker disclosed herein is present. A preferred set of probes may be designed for use in the hybridization assays of the invention in any manner known in the art such that they selectively bind to one allele of a biallelic marker, but not the other under any particular set of assay conditions. Preferred hybridization probes may consists of, consist essentially of, or comprise a contiguous span which ranges in length from 8, 10, 12, 15, 18 or 20 to 25, 35, 40, 50, 60, 70, or 80 nucleotides, or be specified as being 12, 15, 18, 20, 25, 35, 40, or 50 nucleotides in length and including a CNS disorder-related biallelic marker of said sequence. Optionally the original allele or alternative allele disclosed in Table 9 and the first or second allele disclosed in Tables 8 and 10 may be specified as being present at the biallelic marker site. Optionally, said biallelic marker may be within 6, 5, 4, 3, 2, or 1 nucleotides of the center of the hybridization probe or at the center of said probe. A particularly preferred set of hybridization probes is disclosed in Table 14 or a sequence complementary thereto.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances, fluorescent dyes or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes® and others. The solid support is not critical and can be selected by

one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the inventions to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods, which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et

al., Science, 251:767-777, 1991). The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854 and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

Oligonucleotide arrays may comprise at least one of the sequences selected from the group consisting of SEQ ID No. 1-130; and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000, 2000 or 3000 consecutive nucleotides, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for determining whether a sample contains one or more alleles of the biallelic markers of the present invention. Oligonucleotide arrays may also comprise at least one of the sequences selected from the group consisting of SEQ ID No. 1-130; and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000, 2000 or 3000 consecutive nucleotides, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for amplifying one or more alleles of the biallelic markers of Table 7. In other embodiments, arrays may also comprise at least one of the sequences selected from the group consisting of SEQ ID No. 1-130; and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000, 2000 or 3000 consecutive nucleotides, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for conducting microsequencing analyses to determine whether a sample contains one or more alleles of the biallelic markers of the invention. In still further embodiments, the oligonucleotide array may comprise at least one of the sequences selecting from the group consisting of SEQ ID No. 1-130; and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000, 2000 or 3000 nucleotides in length, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for determining whether a sample contains one or more alleles of the biallelic markers of the present invention. In still further embodiments, the oligonucleotide array may comprise at least one of the novel sequences listed in the fifth column of Table 8 or the sequences complementary thereto or a fragment comprising at least 8, 10, 12, 15, 18, 20, 25, 35,

40, 50, 70, 80, 100, 250, 500 or 1000 consecutive nucleotides thereof to the extent that fragments of these lengths are consistent with the lengths of the particular novel sequences.

The present invention also encompasses diagnostic kits comprising one or more polynucleotides of the invention, optionally with a portion or all of the necessary reagents and instructions for genotyping a test subject by determining the identity of a nucleotide at a CNS disorder-related biallelic marker. The determining of the identity may optionally be at a CNS disorder-related biallelic marker that predicts the response of a therapeutic agent, preferably Reboxetine, when administered to a patient suffering from depression. The polynucleotides of a kit may optionally be attached to a solid support, or be part of an array or addressable array of polynucleotides. The kit may provide for the determination of the identity of the nucleotide at a marker position by any method known in the art including, but not limited to, a sequencing assay method, a microsequencing assay method, a hybridization assay method, or an allele specific amplification method. Optionally such a kit may include instructions for scoring the results of the determination with respect to the test subjects' risk of contracting a CNS disorder, or likely response to an agent acting on CNS disorders, or chances of suffering from side effects to an agent acting on CNS disorders.

C. Polypeptides of the Invention

The polynucleotides which encode the WFS1 and the NET polypeptide may include: only the coding sequence for the mature polypeptide; the coding sequence for the polypeptide and additional coding sequence such as a leader or secretory sequence or a proprotein sequence; the coding sequence for the polypeptide (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature polypeptide.

Thus, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence.

As hereinabove indicated, the polynucleotides may have a coding sequence which is a naturally occurring allelic variant of the coding sequence of SEQ ID NO: 543 or 544. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

D. Host Cells

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the WFS1 or NET gene. The culture

conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the E. coli. lac or trp, the phage lambda P.sub.L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in E. coli.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as E. coli, Streptomyces, Salmonella typhimurium; fungal cells, such as yeast; insect cells such as Drosophila S2 and Spodoptera Sf9; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

E. Screening Assays

The polynucleotides of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety f expression vectors for expressing a polypeptide. Such vectors include

chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the E. coli. lac or trp, the phage lambda P.sub.L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as E. coli, Streptomyces, Salmonella typhimurium; fungal cells, such as yeast; insect cells such as Drosophila S2 and Spodoptera Sf9; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3,

pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, PSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are PKK232-8 and PCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P.sub.R, P.sub.L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation: (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples including the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of E. coli and S. cerevisiae TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), .alpha.-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is

assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include E. coli, Bacillus subtilis, Salmonella typhimurium and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, Wis., USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well know to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The WFS1 and NET polypeptides can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The polypeptides of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. Polypeptides of the invention may also include an initial methionine amino acid residue.

F. Screening Assays

The WFS1 protein receptor of the present invention may be employed in a process for screening for antagonists and/or agonists for the receptor.

In general, such screening procedures involve providing appropriate cells which express the receptor on the surface thereof. In particular, a polynucleotide encoding the receptor of the present invention is employed to transfect cells to thereby express the WFS1 receptor. Such transfection may be accomplished by procedures as hereinabove described.

One such screening procedure involves the use of the melanophores which are transfected to express the WFS1 receptor of the present invention. Such a screening technique is described in PCT WO 92/01810 published Feb. 6, 1992.

Thus, for example, such assay may be employed for screening for a receptor antagonist by contacting the melanophore cells which encode the WFS1 receptor with both the receptor ligand and a compound to be screened. Inhibition of the signal generated by the ligand indicates that a compound is a potential antagonist for the receptor, i.e., inhibits activation of the receptor.

The screen may be employed for determining an agonist by contacting such cells with compounds to be screened and determining whether such compound generates a signal, i.e., activates the receptor.

Other screening techniques include the use of cells which express WFS1 receptor (for example, transfected CHO cells) in a system which measures extracellular pH changes caused by receptor activation, for example, as described in Science, volume 246, pages 181-296 (October 1989). For example, potential agonists or antagonists may be contacted with a cell which expresses the WFS1 receptor and a second messenger response, e.g. signal transduction or pH changes, may be measured to determine whether the potential agonist or antagonist is effective.

Another such screening technique involves introducing RNA encoding the WFS1 receptor into xenopus oocytes to transiently express the receptor. The receptor oocytes may then be contacted in the case of antagonist screening with the receptor ligand and a compound to be screened, followed by detection of inhibition of a calcium signal.

Another screening technique involves expressing the WFS1 receptor in which the receptor is linked to a phospholipase C or D. As representative examples of such cells, there may be mentioned endothelial cells, smooth muscle cells, embryonic kidney cells, etc. The screening for an antagonist or agonist may be accomplished as hereinabove described by detecting activation of the receptor or inhibition of activation of the receptor from the phospholipase second signal.

Another method involves screening for antagonists by determining inhibition of binding of labeled ligand to cells which have the receptor on the surface thereof. Such a method involves transfecting a eukaryotic cell with DNA encoding the WFS1 receptor such that the cell expresses the receptor on its surface and contacting the cell with a potential antagonist in the presence of a labeled form of a known ligand. The ligand can be labeled, e.g., by radioactivity. The amount of labeled ligand bound to the receptors is measured, e.g., by measuring radioactivity of the receptors. If the potential antagonist binds to the receptor as determined by a reduction of labeled ligand which binds to the receptors, the binding of labeled ligand to the receptor is inhibited.

The present invention also provides a method for determining whether a ligand not known to be capable of binding to a WFS1 receptor can bind to such receptor which comprises contacting a mammalian cell which expresses a WFS1 receptor with the ligand under conditions permitting binding of ligands to the WFS1 receptor, detecting the presence of a ligand which binds to the receptor and thereby determining whether the ligand binds to the WFS1 receptor. The systems hereinabove described for determining agonists and/or antagonists may also be employed for determining ligands which bind to the receptor.

In general, antagonists for WFS1 receptors which are determined by screening procedures may be employed for a variety of therapeutic purposes. For example, such antagonists have been employed for treatment of hypertension, angina pectoris, myocardial infarction, ulcers, asthma, allergies, psychoses, depression, migraine, vomiting, stroke, eating disorders, migraine headaches, cancer and benign prostatic hypertrophy.

Agonists for WFS1 receptors are also useful for therapeutic purposes, such as the treatment of Wolfram syndrome and/or depression.

Examples of WFS1 receptor antagonists include an antibody, or in some cases an oligonucleotide, which binds to the WFS1 receptor but does not elicit a second messenger response such that the activity of the WFS1 receptor is prevented. Antibodies include anti-idiotypic antibodies which recognize unique determinants generally associated with the antigen-binding site of an antibody.

Potential antagonists also include proteins which are closely related to the ligand of the WFS1 receptor, i.e. a fragment of the ligand, which have lost biological function and when binding to the WFS1 receptor, elicit no response.

A potential antagonist also includes an antisense construct prepared through the use of antisense technology. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide to DNA or RNA. For example, the 5' coding portion of the polynucleotide sequence, which encodes for the mature polypeptides of the present invention, is used to design an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription (triple helix -see Lee et al., Nucl. Acids Res., 6:3073 (1979); Cooney et al, Science, 241:456 (1988); and Dervan et al., Science, 251: 1360 (1991)), thereby preventing transcription and the production of WFS1 receptor. The antisense RNA oligonucleotide hybridizes to the mRNA in vivo and blocks translation of the mRNA molecule into the WFS1 receptor (antisense--Okano, J. Neurochem., 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, Fla. (1988)). The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed in vivo to inhibit production of WFS1 receptor.

Another potential antagonist is a small molecule which binds to the WFS1 receptor, making it inaccessible to ligands such that normal biological activity is prevented. Examples of small molecules include but are not limited to small peptides or peptide-like molecules.

Potential antagonists also include a soluble form of a WFS1 receptor, e.g. a fragment of the receptor, which binds to the ligand and prevents the ligand from interacting with membrane bound WFS1 receptors.

The WFS1 receptor and antagonists or agonists may be employed in combination with a suitable pharmaceutical carrier. Such compositions comprise a therapeutically effective amount of the polypeptide, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation should suit the mode of administration.

G. Antibodies

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

IV. Methods for *De Novo* Identification of Biallelic Markers

Large fragments of human DNA, carrying genes of interest involved in CNS disorders; were cloned, sequenced and screened for biallelic markers. Biallelic markers within the candidate genes themselves as well as markers located on the same genomic fragment were identified. It will be clear to one of skill in the art that large fragments of human genomic DNA may be obtained from any appropriate source and may be cloned into a number of suitable vectors.

In a preferred embodiment of the invention, BAC (Bacterial Artificial Chromosomes) vectors were used to construct DNA libraries covering the entire human genome. Specific amplification primers were designed for each candidate gene and the BAC library was screened by PCR until there was at least one positive BAC clone per candidate gene. Genomic sequence, screened for biallelic markers, was generated by sequencing ends of BAC subclones. Details of a preferred embodiment are provided in Example 1. As a preferred alternative to sequencing the ends of an adequate number of BAC subclones, high throughput deletion-based sequencing vectors, which allow the generation of a high quality sequence information covering fragments of about 6kb, may be used. Having sequence fragments longer than 2.5 or 3kb enhances the chances of identifying biallelic markers therein. Methods of constructing and sequencing a nested set of deletions are disclosed in the related U.S. Patent Application entitled "High Throughput DNA Sequencing Vector" (Serial No. 09/058,746).

In another embodiment of the invention, genomic sequences of candidate genes were available in public databases allowing direct screening for biallelic markers.

Any of a variety of methods can be used to screen a genomic fragment for single nucleotide polymorphisms such as differential hybridization with oligonucleotide probes, detection of changes in the mobility measured by gel electrophoresis or direct sequencing of the amplified nucleic acid. A preferred method for identifying biallelic markers involves comparative sequencing of genomic DNA fragments from an appropriate number of unrelated individuals.

In a first embodiment, DNA samples from unrelated individuals are pooled together, following which the genomic DNA of interest is amplified and sequenced. The nucleotide sequences thus obtained are then analyzed to identify significant polymorphisms. One of the

major advantages of this method resides in the fact that the pooling of the DNA samples substantially reduces the number of DNA amplification reactions and sequencing reactions, which must be carried out. Moreover, this method is sufficiently sensitive so that a biallelic marker obtained thereby usually demonstrates a sufficient frequency of its less common allele to be useful in conducting association studies. Usually, the frequency of the least common allele of a biallelic marker identified by this method is at least 10%.

In a second embodiment, the DNA samples are not pooled and are therefore amplified and sequenced individually. This method is usually preferred when biallelic markers need to be identified in order to perform association studies within candidate genes. Preferably, highly relevant gene regions such as promoter regions or exon regions may be screened for biallelic markers. A biallelic marker obtained using this method may show a lower degree of informativeness for conducting association studies, e.g. if the frequency of its less frequent allele may be less than about 10%. Such a biallelic marker will however be sufficiently informative to conduct association studies and it will further be appreciated that including less informative biallelic markers in the genetic analysis studies of the present invention, may allow in some cases the direct identification of causal mutations, which may, depending on their penetrance, be rare mutations.

The following is a description of the various parameters of a preferred method used by the inventors for the identification of the biallelic markers of the present invention.

A. Genomic DNA Samples

The genomic DNA samples from which the biallelic markers of the present invention are generated are preferably obtained from unrelated individuals corresponding to a heterogeneous population of known ethnic background. The number of individuals from whom DNA samples are obtained can vary substantially, preferably from about 10 to about 1000, more preferably from about 50 to about 200 individuals. Usually, DNA samples are collected from at least about 100 individuals in order to have sufficient polymorphic diversity in a given population to identify as many markers as possible and to generate statistically significant results.

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples, which can be tested by the methods of the present invention described herein, and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the present invention is from peripheral venous blood of each donor. Techniques to prepare genomic DNA from biological samples are well known to the skilled technician. Details

of a preferred embodiment are provided in Example 1. The person skilled in the art can choose to amplify pooled or unpooled DNA samples.

B. DNA Amplification

The identification of biallelic markers in a sample of genomic DNA may be facilitated through the use of DNA amplification methods. DNA samples can be pooled or unpooled for the amplification step. DNA amplification techniques are well known to those skilled in the art. Various methods to amplify DNA fragments carrying biallelic markers are further described herein. The PCR technology is the preferred amplification technique used to identify new biallelic markers.

In a first embodiment, biallelic markers are identified using genomic sequence information generated by the inventors. Genomic DNA fragments, such as the inserts of the BAC clones described above, are sequenced and used to design primers for the amplification of 500 bp fragments. These 500 bp fragments are amplified from genomic DNA and are scanned for biallelic markers. Primers may be designed using the OSP software (Hillier L. and Green P., 1991). All primers may contain, upstream of the specific target bases, a common oligonucleotide tail that serves as a sequencing primer. Those skilled in the art are familiar with primer extensions, which can be used for these purposes.

In another embodiment of the invention, genomic sequences of candidate genes are available in public databases allowing direct screening for biallelic markers. Preferred primers, useful for the amplification of genomic sequences encoding the candidate genes, focus on promoters, exons and splice sites of the genes. A biallelic marker present in these functional regions of the gene has a higher probability to be a causal mutation.

Preferred primers include those disclosed in Table 13.

C. Sequencing of Amplified Genomic DNA and Identification of Single Nucleotide Polymorphisms

The amplification products generated as described above, are then sequenced using any method known and available to the skilled technician. Methods for sequencing DNA using either the dideoxy-mediated method (Sanger method) or the Maxam-Gilbert method are widely known to those of ordinary skill in the art. Such methods are for example disclosed in Maniatis et al. (Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Second Edition, 1989). Alternative approaches include hybridization to high-density DNA probe arrays as described in Chee et al. (Science 274, 610, 1996).

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. The products of the sequencing reactions are run on sequencing gels and the sequences are determined using gel image analysis. The polymorphism search is based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position. Because each dideoxy

terminator is labeled with a different fluorescent molecule, the two peaks corresponding to a biallelic site present distinct colors corresponding to two different nucleotides at the same position on the sequence. However, the presence of two peaks can be an artifact due to background noise. To exclude such an artifact, the two DNA strands are sequenced and a comparison between the peaks is carried out. In order to be registered as a polymorphic sequence, the polymorphism has to be detected on both strands.

The above procedure permits those amplification products, which contain biallelic markers to be identified. The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is approximately 0.1 for the minor allele, as verified by sequencing pools of known allelic frequencies. However, more than 90% of the biallelic polymorphisms detected by the pooling method have a frequency for the minor allele higher than 0.25. Therefore, the biallelic markers selected by this method have a frequency of at least 0.1 for the minor allele and less than 0.9 for the major allele. Preferably at least 0.2 for the minor allele and less than 0.8 for the major allele, more preferably at least 0.3 for the minor allele and less than 0.7 for the major allele, thus a heterozygosity rate higher than 0.18, preferably higher than 0.32, more preferably higher than 0.42.

In another embodiment, biallelic markers are detected by sequencing individual DNA samples; the frequency of the minor allele of such a biallelic marker may be less than 0.1.

The markers carried by the same fragment of genomic DNA, such as the insert in a BAC clone, need not necessarily be ordered with respect to one another within the genomic fragment to conduct association studies. However, in some embodiments of the present invention, the order of biallelic markers carried by the same fragment of genomic DNA are determined.

D. Validation of the Biallelic Markers of the Present Invention

The polymorphisms are evaluated for their usefulness as genetic markers by validating that both alleles are present in a population. Validation of the biallelic markers is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. Microsequencing is a preferred method of genotyping alleles. The validation by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group can be as small as one individual if that individual is heterozygous for the allele in question. Preferably the group contains at least three individuals, more preferably the group contains five or six individuals, so that a single validation test will be more likely to result in the validation of more of the biallelic markers that are being tested. It should be noted, however, that when the validation test is performed on a small group it may result in a false negative result if as a result of sampling error none of the individuals tested carries one of the two alleles. Thus, the validation process is less useful in demonstrating that a particular initial result is an artifact, than it is at demonstrating that there is a *bona fide* biallelic marker at a particular position in a

sequence. For an indication of whether a particular biallelic marker has been validated see Table 7. All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with validated biallelic markers.

E. Evaluation of the Frequency of the Biallelic Markers of the Present Invention

The validated biallelic markers are further evaluated for their usefulness as genetic markers by determining the frequency of the least common allele at the biallelic marker site. The determination of the least common allele is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. This determination of frequency by genotyping step may be performed on individual samples derived from each individual ~~in the~~ group or by genotyping a pooled sample derived from more than one individual. The group must be large enough to be representative of the population as a whole. Preferably the group contains at least 20 individuals, more preferably the group contains at least 50 individuals, most preferably the group contains at least 100 individuals. Of course the larger the group the greater the accuracy of the frequency determination because of reduced sampling error. For an indication of the frequency for the less common allele of a particular biallelic marker of the invention see Table 7. A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker." All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with high quality biallelic markers.

V. Methods of Genotyping an Individual for Biallelic Markers

Methods are provided to genotype a biological sample for one or more biallelic markers of the present invention, all of which may be performed *in vitro*. Such methods of genotyping comprise determining the identity of a nucleotide at a CNS disorder-related biallelic marker by any method known in the art. These methods find use in genotyping case-control populations in association studies as well as individuals in the context of detection of alleles of biallelic markers which, are known to be associated with a given trait, in which case both copies of the biallelic marker present in individual's genome are determined so that an individual may be classified as homozygous or heterozygous for a particular allele.

These genotyping methods can be performed nucleic acid samples derived from a single individual or pooled DNA samples.

Genotyping can be performed using similar methods as those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic fragments from different individuals is used to identify new biallelic markers whereas microsequencing is used for genotyping known biallelic markers in diagnostic and association study applications.

A. Source of DNA for Genotyping

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA may be extracted from cells, tissues, body fluids and the like as described herein. While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are taken are generally understood to be human.

B. Amplification of DNA Fragments Comprising Biallelic Markers

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers may be used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention.

Amplification of DNA may be achieved by any method known in the art. The established PCR (polymerase chain reaction) method or by developments thereof or alternatives. Amplification methods which can be utilized herein include but are not limited to Ligase Chain Reaction (LCR) as described in EP A 320 308 and EP A 439 182, Gap LCR (Wolcott, M.J., *Clin. Microbiol. Rev.* 5:370-386), the so-called "NASBA" or "3SR" technique described in Guatelli J.C. et al. (*Proc. Natl. Acad. Sci. USA* 87:1874-1878, 1990) and in Compton J. (*Nature* 350:91-92, 1991), Q-beta amplification as described in European Patent Application no 4544610, strand displacement amplification as described in Walker et al. (*Clin. Chem.* 42:9-13, 1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the

ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall R.L. et al. (*PCR Methods and Applications* 4:80-84, 1994). AGLCR is a modification of GLCR that allows the amplification of RNA.

Some of these amplification methods are particularly suited for the detection of single nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as it is further described herein.

The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering White, B.A. Ed. in *Methods in Molecular Biology* 67: Humana Press, Totowa (1997) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patents 4,683,195, 4,683,202 and 4,965,188.

The identification of biallelic markers as described above allows the design of appropriate oligonucleotides, which can be used as primers to amplify DNA fragments comprising the biallelic markers of the present invention. Amplification can be performed using the primers initially used to discover new biallelic markers which are described herein or any set of primers allowing the amplification of a DNA fragment comprising a biallelic marker of the present invention. Primers can be prepared by any suitable method. As for example, direct chemical synthesis by a method such as the phosphodiester method of Narang S.A. et al. (*Methods Enzymol.* 68:90-98, 1979), the phosphodiester method of Brown E.L. et al. (*Methods*

Enzymol. 68:109-151, 1979), the diethylphosphoramidite method of Beaucage et al. (*Tetrahedron Lett.* 22:1859-1862, 1981) and the solid support method described in EP 0 707 592.

In some embodiments the present invention provides primers for amplifying a DNA fragment containing one or more biallelic markers of the present invention. Preferred amplification primers are listed in Table 13. It will be appreciated that the primers listed are merely exemplary and that any other set of primers which produce amplification products containing one or more biallelic markers of the present invention.

The primers are selected to be substantially complementary to the different strands of each specific sequence to be amplified. The length of the primers of the present invention can range from 8 to 100 nucleotides, preferably from 8 to 50, 8 to 30 or more preferably 8 to 25 nucleotides. Shorter primers tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The formation of stable hybrids depends on the melting temperature (Tm) of the DNA. The Tm depends on the length of the primer, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer, the higher is the melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The G+C content of the amplification primers of the present invention preferably ranges between 10 and 75 %, more preferably between 35 and 60 %, and most preferably between 40 and 55 %. The appropriate length for primers under a particular set of assay conditions may be empirically determined by one of skill in the art.

The spacing of the primers determines the length of the segment to be amplified. In the context of the present invention amplified segments carrying biallelic markers can range in size from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. It will be appreciated that amplification primers for the biallelic markers may be any sequence which allow the specific amplification of any DNA fragment carrying the markers. Amplification primers may be labeled or immobilized on a solid support as described in I.

C. Methods of Genotyping DNA samples for Biallelic Markers

Any method known in the art can be used to identify the nucleotide present at a biallelic marker site. Since the biallelic marker allele to be detected has been identified and specified in the present invention, detection will prove simple for one of ordinary skill in the art by employing any of a number of techniques. Many genotyping methods require the previous amplification of the DNA region carrying the biallelic marker of interest. While the amplification of target or signal is often preferred at present, ultrasensitive detection methods which do not require amplification are also encompassed by the present genotyping methods. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms

include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al. (*Proc. Natl. Acad. Sci. U.S.A* 86:27776-2770, 1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield, V.C. et al. (*Proc. Natl. Acad. Sci. USA* 49:699-706, 1991), White et al. (*Genomics* 12:301-306, 1992), Grompe, M. et al. (*Proc. Natl. Acad. Sci. USA* 86:5855-5892, 1989) and Grompe, M. (*Nature Genetics* 5:111-117, 1993). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US patent 4,656,127.

Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by sequencing assay, enzyme-based mismatch detection assay, or hybridization assay. The following is a description of some preferred methods. A highly preferred method is the microsequencing technique. The term "sequencing assay" is used herein to refer to polymerase extension of duplex primer/template complexes and includes both traditional sequencing and microsequencing.

i) Sequencing assays

The nucleotide present at a polymorphic site can be determined by sequencing methods. In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as described above. DNA sequencing methods are described herein.

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Sequence analysis allows the identification of the base present at the biallelic marker site.

ii) Microsequencing assays

In microsequencing methods, a nucleotide at the polymorphic site that is unique to one of the alleles in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of a polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the selected nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way.

Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883. Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously. An example of a typical microsequencing procedure that can be used in the context of the present invention is provided in Example 2.

Different approaches can be used to detect the nucleotide added to the microsequencing primer. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (*Nucleic Acids Research* 25:347-353 1997) and Chen et al. (*Proc. Natl. Acad. Sci. USA* 94/20 10756-10761,1997). In this method amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein–labeled primer in the presence of allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time. Alternatively, the extended primer may be analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff L.A. and Smirnov I.P., *Genome Research*, 7:378-388, 1997).

Microsequencing may be achieved by the established microsequencing method or by developments or derivatives thereof. Alternative methods include several solid-phase microsequencing techniques. The basic microsequencing protocol is the same as described previously, except that the method is conducted as a heterogenous phase assay, in which the primer or the target molecule is immobilized or captured onto a solid support. To simplify the primer separation and the terminal nucleotide addition analysis, oligonucleotides are attached to solid supports or are modified in such ways that permit affinity separation as well as polymerase extension. The 5' ends and internal nucleotides of synthetic oligonucleotides can be modified in a number of different ways to permit different affinity separation approaches, e.g., biotinylation. If a single affinity group is used on the oligonucleotides, the oligonucleotides can be separated from the incorporated terminator reagent. This eliminates the need of physical or size separation. More than one oligonucleotide can be separated from the terminator reagent and analyzed simultaneously if more than one affinity group is used. This permits the analysis of several nucleic acid species or more nucleic acid sequence information per extension reaction. The affinity group need not be on the priming oligonucleotide but could alternatively be present on the template. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles. In the same manner oligonucleotides or templates may be attached to a solid support in a high-density format. In such solid phase microsequencing reactions, incorporated ddNTPs can be radiolabeled (Syvänen, *Clinica Chimica Acta* 226:225-236, 1994) or linked to fluorescein (Livak and Hainer, *Human Mutation* 3:379-385,1994). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of antifluorescein antibody conjugated with alkaline phosphatase, followed by incubation with a chromogenic substrate (such as *p*-nitrophenyl

phosphate). Other possible reporter-detection pairs include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (Harju et al., *Clin. Chem.* 39/11 2282-2287, 1993) or biotinylated ddNTP and horseradish peroxidase-conjugated streptavidin with *o*-phenylenediamine as a substrate (WO 92/15712). As yet another alternative solid-phase microsequencing procedure, Nyren et al. (*Analytical Biochemistry* 208:171-175, 1993) described a method relying on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA).

Pastinen et al. (*Genome research* 7:606-614, 1997) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described herein.

In one aspect the present invention provides polynucleotides and methods to genotype one or more biallelic markers of the present invention by performing a microsequencing assay. Preferred microsequencing primers include those being featured in Table 12. It will be appreciated that the microsequencing primers listed in Table 12 are merely exemplary and that, any primer having a 3' end immediately adjacent to a polymorphic nucleotide may be used. Similarly, it will be appreciated that microsequencing analysis may be performed for any biallelic marker or any combination of biallelic markers of the present invention. One aspect of the present invention is a solid support which includes one or more microsequencing primers listed in Table 12, or fragments comprising at least 8, at least 12, at least 15, or at least 20 consecutive nucleotides thereof and having a 3' terminus immediately upstream of the corresponding biallelic marker, for determining the identity of a nucleotide at a biallelic marker site.

iii) Mismatch detection assays based on polymerases and ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by mismatch detection assays based on polymerases and/or ligases. These assays are based on the specificity of polymerases and ligases. Polymerization reactions places particularly stringent requirements on correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end. The terms "enzyme based mismatch detection assay" are used herein to refer to any method of determining the allele of a biallelic marker based on the specificity of ligases and polymerases. Preferred methods are described below. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described herein.

1. Allele specific amplification

Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy, whereby one of the alleles is amplified without amplification of the other allele. This is accomplished by placing a polymorphic base at the 3' end of one of the amplification primers. Because the extension forms from the 3'end of the primer, a mismatch at or near this position has an inhibitory effect on amplification. Therefore, under appropriate amplification conditions, these primers only direct amplification on their complementary allele. Designing the appropriate allele-specific primer and the corresponding assay conditions are well with the ordinary skill in the art.

2. Ligation/amplification based methods

The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable of detecting biallelic markers and may be advantageously combined with PCR as described by Nickerson D.A. et al. (*Proc. Natl. Acad. Sci. U.S.A.* 87:8923-8927, 1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other methods which are particularly suited for the detection of biallelic markers include LCR (ligase chain reaction), Gap LCR (GLCR) which are described herein. As mentioned above LCR uses two pairs of probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides, is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide(s) that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when they hybridize to the target molecule, a "gap" is created as described in WO 90/01069. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

Ligase/Polymerase-mediated Genetic Bit Analysis™ is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271). This

method involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

iv) Hybridization assay methods

A preferred method of determining the identity of the nucleotide present at a biallelic marker site involves nucleic acid hybridization. The hybridization probes, which can be conveniently used in such reactions, preferably include the probes defined herein. Any hybridization assay may be used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., Molecular Cloning – A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989).

Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions, under which a probe will hybridize only to the exactly complementary target sequence are well known in the art (Sambrook et al., Molecular Cloning – A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989). Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Sodium citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2

x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. By way of example and not limitation, procedures using conditions of intermediate stringency are as follows: Filters containing DNA are prehybridized, and then hybridized at a temperature of 60°C in the presence of a 5 x SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2x SSC at 50°C and the hybridized probes are detectable by autoradiography. Other conditions of high and intermediate stringency which may be used are well known in the art and as cited in Sambrook et al. (Molecular Cloning - A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989) and Ausubel et al. (Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y., 1989).

Although such hybridizations can be performed in solution, it is preferred to employ a solid-phase hybridization assay. The target DNA comprising a biallelic marker of the present invention may be amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps may be employed to wash away excess target DNA or probe. Standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes.

Two recently developed assays allow hybridization-based allele discrimination with no need for separations or washes (see Landegren U. et al., *Genome Research*, 8:769-776,1998). The TaqMan assay takes advantage of the 5' nuclease activity of Taq DNA polymerase to digest a DNA probe annealed specifically to the accumulating amplification product. TaqMan probes are labeled with a donor-acceptor dye pair that interacts via fluorescence energy transfer. Cleavage of the TaqMan probe by the advancing polymerase during amplification dissociates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence. All reagents necessary to detect two allelic variants can be assembled at the beginning of the reaction and the results are monitored in real time (see Livak et al., *Nature Genetics*, 9:341-342, 1995). In an alternative homogeneous hybridization-based procedure, molecular beacons are used for allele discriminations. Molecular beacons are hairpin-shaped oligonucleotide probes that report the presence of specific nucleic acids in homogeneous solutions. When they bind to their targets they undergo a conformational reorganization that restores the fluorescence of an internally quenched fluorophore (Tyagi et al., *Nature Biotechnology*, 16:49-53, 1998).

The polynucleotides provided herein can be used in hybridization assays for the detection of biallelic marker alleles in biological samples. These probes are characterized in that they preferably comprise between 8 and 50 nucleotides, and in that they are sufficiently complementary to a sequence comprising a biallelic marker of the present invention to hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation. The GC content in the probes of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %. The length of these probes can range from 10, 15, 20, or 30 to at least 100 nucleotides, preferably from 10 to 50, more preferably from 18 to 35 nucleotides. A particularly preferred probe is 25 nucleotides in length. Preferably the biallelic marker is within 4 nucleotides of the center of the polynucleotide probe. In particularly preferred probes the biallelic marker is at the center of said polynucleotide. Shorter probes may lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes are expensive to produce and can sometimes self-hybridize to form hairpin structures. Methods for the synthesis of oligonucleotide probes have been described above and can be applied to the probes of the present invention.

Preferably the probes of the present invention are labeled or immobilized on a solid support. Labels and solid supports are further described in I. Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702; morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified, U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993 describes modifications, which can be used to render a probe non-extendable.

The probes of the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA or Northern hybridization to mRNA. The probes can also be used to detect PCR amplification products. By assaying the hybridization to an allele specific probe, one can detect the presence or absence of a biallelic marker allele in a given sample.

High-Throughput parallel hybridizations in array format are specifically encompassed within "hybridization assays" and are described below.

i. Hybridization to addressable arrays of oligonucleotides

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (the chip) at selected positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (Hacia et al., *Nature Genetics*, 14(4):441-447, 1996; Shoemaker et al., *Nature Genetics*, 14(4):450-456, 1996 ; Kozal et al., *Nature Medicine*, 2:753-759, 1996). Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip™), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP785280 describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be "tiled" for a large number of specific polymorphisms. By "tiling" is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of monomers, i.e. nucleotides. Tiling strategies are further described in PCT application No. WO 95/11995. In a particular aspect, arrays are tiled for a number of specific, identified biallelic marker sequences. In particular the array is tiled to include a number of detection blocks, each detection block being specific for a specific biallelic marker or a set of biallelic markers. For example, a detection block may be tiled to include a number of probes, which span the sequence segment that includes a specific polymorphism. To ensure probes that are complementary to each allele, the probes are synthesized in pairs differing at the biallelic marker. In addition to the probes differing at the polymorphic base, monosubstituted probes are also generally tiled within the detection block. These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the biallelic marker. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artefactual cross-hybridization. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is

then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning may be carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No. 5,424,186.

Thus, in some embodiments, the chips may comprise an array of nucleic acid sequences of fragments of about 15 nucleotides in length. In further embodiments, the chip may comprise an array including at least one of the sequences selected from the group consisting of SEQ ID No. 1-130 and the sequences complementary thereto, or a fragment thereof at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides. In some embodiments, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more of these polynucleotides of the invention. Solid supports and polynucleotides of the present invention attached to solid supports are further described in I.

v) Integrated systems

Another technique, which may be used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip. For genotyping biallelic markers, the microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

VI. Methods of Genetic Analysis Using the Biallelic Markers of the Present Invention

Different methods are available for the genetic analysis of complex traits (see Lander and Schork, *Science*, 265, 2037-2048, 1994). The search for disease-susceptibility genes is conducted using two main methods: the linkage approach in which evidence is sought for cosegregation between a locus and a putative trait locus using family studies, and the association approach in which evidence is sought for a statistically significant association between an allele and a trait or a trait causing allele (Khoury J. et al., *Fundamentals of Genetic Epidemiology, Oxford University Press, NY*, 1993). In general, the biallelic markers of the present invention find use in any method known in the art to demonstrate a statistically significant correlation between a genotype and a phenotype. The biallelic markers may be used in parametric and non-parametric linkage analysis methods. Preferably, the biallelic markers of the present invention are used to identify genes associated with detectable traits using association studies, an approach

which does not require the use of affected families and which permits the identification of genes associated with complex and sporadic traits.

The genetic analysis using the biallelic markers of the present invention may be conducted on any scale. The whole set of biallelic markers of the present invention or any subset of biallelic markers of the present invention may be used. In some embodiments a subset of biallelic markers corresponding to one or several candidate genes of the present invention may be used. In other embodiments a subset of biallelic markers corresponding to CNS disorder candidate genes may be used. Alternatively, a subset of biallelic markers of the present invention localised on a specific chromosome segment may be used. Further, any set of genetic markers including a biallelic marker of the present invention may be used. A set of biallelic polymorphisms that, could be used as genetic markers in combination with the biallelic markers of the present invention, has been described in WO 98/20165. As mentioned above, it should be noted that the biallelic markers of the present invention may be included in any complete or partial genetic map of the human genome. These different uses are specifically contemplated in the present invention and claims.

A. Linkage Analysis

Linkage analysis is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. Thus, the aim of linkage analysis is to detect marker loci that show cosegregation with a trait of interest in pedigrees.

i. Parametric methods

When data are available from successive generations there is the opportunity to study the degree of linkage between pairs of loci. Estimates of the recombination fraction enable loci to be ordered and placed onto a genetic map. With loci that are genetic markers, a genetic map can be established, and then the strength of linkage between markers and traits can be calculated and used to indicate the relative positions of markers and genes affecting those traits (Weir, B.S., *Genetic data Analysis II: Methods for Discrete population genetic Data, Sinauer Assoc., Inc., Sunderland, MA, USA,* 1996). The classical method for linkage analysis is the logarithm of odds (lod) score method (see Morton N.E., *Am.J. Hum.Genet.*, 7:277-318, 1955; Ott J., *Analysis of Human Genetic Linkage, John Hopkins University Press, Baltimore,* 1991). Calculation of lod scores requires specification of the mode of inheritance for the disease (parametric method). Generally, the length of the candidate region identified using linkage analysis is between 2 and 20Mb. Once a candidate region is identified as described above, analysis of recombinant individuals using additional markers allows further delineation of the candidate region. Linkage analysis studies have generally relied on the use of a maximum of 5,000 microsatellite markers, thus limiting the maximum theoretical attainable resolution of linkage analysis to about 600 kb on average.

Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns and which have a high penetrance (i.e., the ratio between the number of trait positive carriers of allele *a* and the total number of *a* carriers in the population). However, parametric linkage analysis suffers from a variety of drawbacks. First, it is limited by its reliance on the choice of a genetic model suitable for each studied trait. Furthermore, as already mentioned, the resolution attainable using linkage analysis is limited, and complementary studies are required to refine the analysis of the typical 2Mb to 20Mb regions initially identified through linkage analysis. In addition, parametric linkage analysis approaches have proven difficult when applied to complex genetic traits, such as those due to the combined action of multiple genes and/or environmental factors. It is very difficult to model these factors adequately in a lod score analysis. In such cases, too large an effort and cost are needed to recruit the adequate number of affected families required for applying linkage analysis to these situations, as recently discussed by Risch, N. and Merikangas, K. (*Science*, 273:1516-1517, 1996).

ii. Non-parametric methods

The advantage of the so-called non-parametric methods for linkage analysis is that they do not require specification of the mode of inheritance for the disease, they tend to be more useful for the analysis of complex traits. In non-parametric methods, one tries to prove that the inheritance pattern of a chromosomal region is not consistent with random Mendelian segregation by showing that affected relatives inherit identical copies of the region more often than expected by chance. Affected relatives should show excess "allele sharing" even in the presence of incomplete penetrance and polygenic inheritance. In non-parametric linkage analysis the degree of agreement at a marker locus in two individuals can be measured either by the number of alleles identical by state (IBS) or by the number of alleles identical by descent (IBD). Affected sib pair analysis is a well-known special case and is the simplest form of these methods.

The biallelic markers of the present invention may be used in both parametric and non-parametric linkage analysis. Preferably biallelic markers may be used in non-parametric methods which allow the mapping of genes involved in complex traits. The biallelic markers of the present invention may be used in both IBD- and IBS- methods to map genes affecting a complex trait. In such studies, taking advantage of the high density of biallelic markers, several adjacent biallelic marker loci may be pooled to achieve the efficiency attained by multi-allelic markers (Zhao et al., *Am. J. Hum. Genet.*, 63:225-240, 1998).

However, both parametric and non-parametric linkage analysis methods analyse affected relatives, they tend to be of limited value in the genetic analysis of drug responses or in the analysis of side effects to treatments. This type of analysis is impractical in such cases due to the lack of availability of familial cases. In fact, the likelihood of having more than one individual in a family being exposed to the same drug at the same time is extremely low.

B. Population Association Studies

The present invention comprises methods for identifying one or several genes among a set of candidate genes that are associated with a detectable trait using the biallelic markers of the present invention. In one embodiment the present invention comprises methods to detect an association between a biallelic marker allele or a biallelic marker haplotype and a trait. Further, the invention comprises methods to identify a trait causing allele in linkage disequilibrium with any biallelic marker allele of the present invention.

As described above, alternative approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. In a preferred embodiment, the biallelic markers of the present invention are used to perform candidate gene association studies. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait when some information concerning the biology of the trait is available. Further, the biallelic markers of the present invention may be incorporated in any map of genetic markers of the human genome in order to perform genome-wide association studies. Methods to generate a high-density map of biallelic markers has been described in US Provisional Patent application serial number 60/082,614. The biallelic markers of the present invention may further be incorporated in any map of a specific candidate region of the genome (a specific chromosome or a specific chromosomal segment for example).

As mentioned above, association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families. Association studies are extremely valuable as they permit the analysis of sporadic or multifactor traits. Moreover, association studies represent a powerful method for fine-scale mapping enabling much finer mapping of trait causing alleles than linkage studies. Studies based on pedigrees often only narrow the location of the trait causing allele. Association studies using the biallelic markers of the present invention can therefore be used to refine the location of a trait causing allele in a candidate region identified by Linkage Analysis methods. Moreover, once a chromosome segment of interest has been identified, the presence of a candidate gene such as a candidate gene of the present invention, in the region of interest can provide a shortcut to the identification of the trait causing allele. Biallelic markers of the present invention can be used to demonstrate that a candidate gene is associated with a trait. Such uses are specifically contemplated in the present invention and claims.

i. Determining the frequency of a biallelic marker allele or of a biallelic marker haplotype in a population

Association studies explore the relationships among frequencies for sets of alleles between loci.

1) Determining the frequency of an allele in a population

Allelic frequencies of the biallelic markers in a population can be determined using one of the methods described above under the heading "Methods for genotyping an individual for biallelic markers", or any genotyping procedure suitable for this intended purpose. Genotyping pooled samples or individual samples can determine the frequency of a biallelic marker allele in a population. One way to reduce the number of genotypings required is to use pooled samples. A major obstacle in using pooled samples is in terms of accuracy and reproducibility for determining accurate DNA concentrations in setting up the pools. Genotyping individual samples provides higher sensitivity, reproducibility and accuracy and; is the preferred method used in the present invention. Preferably, each individual is genotyped separately and simple gene counting is applied to determine the frequency of ~~an~~ allele of a biallelic marker or of a genotype in a given population.

2) Determining the frequency of a haplotype in a population

The gametic phase of haplotypes is unknown when diploid individuals are heterozygous at more than one locus. Using genealogical information in families gametic phase can sometimes be inferred (Perlin et al., *Am. J. Hum. Genet.*, 55:777-787, 1994). When no genealogical information is available different strategies may be used. One possibility is that the multiple-site heterozygous diploids can be eliminated from the analysis, keeping only the homozygotes and the single-site heterozygote individuals, but this approach might lead to a possible bias in the sample composition and the underestimation of low-frequency haplotypes. Another possibility is that single chromosomes can be studied independently, for example, by asymmetric PCR amplification (see Newton et al., *Nucleic Acids Res.*, 17:2503-2516, 1989; Wu et al., *Proc. Natl. Acad. Sci. USA*, 86:2757, 1989) or by isolation of single chromosome by limit dilution followed by PCR amplification (see Ruano et al., *Proc. Natl. Acad. Sci. USA*, 87:6296-6300, 1990). Further, a sample may be haplotyped for sufficiently close biallelic markers by double PCR amplification of specific alleles (Sarkar, G. and Sommer S.S., *Biotechniques*, 1991). These approaches are not entirely satisfying either because of their technical complexity, the additional cost they entail, their lack of generalisation at a large scale, or the possible biases they introduce. To overcome these difficulties, an algorithm to infer the phase of PCR-amplified DNA genotypes introduced by Clark A.G. (*Mol. Biol. Evol.*, 7:111-122, 1990) may be used. Briefly, the principle is to start filling a preliminary list of haplotypes present in the sample by examining unambiguous individuals, that is, the complete homozygotes and the single-site heterozygotes. Then other individuals in the same sample are screened for the possible occurrence of previously recognized haplotypes. For each positive identification, the complementary haplotype is added to the list of recognized haplotypes, until the phase information for all individuals is either resolved or identified as unresolved. This method assigns a single haplotype to each multiheterozygous individual, whereas several haplotypes are possible when there are more than one heterozygous site. Alternatively, one can use methods estimating haplotype frequencies in a population without

assigning haplotypes to each individual. Preferably, a method based on an expectation-maximization (EM) algorithm (Dempster et al., *J. R. Stat. Soc.*, 39B: 1-38, 1977) leading to maximum-likelihood estimates of haplotype frequencies under the assumption of Hardy-Weinberg proportions (random mating) is used (see Excoffier L. and Slatkin M., *Mol. Biol. Evol.*, 12(5): 921-927, 1995). The EM algorithm is a generalized iterative maximum-likelihood approach to estimation that is useful when data are ambiguous and/or incomplete. The EM algorithm is used to resolve heterozygotes into haplotypes. Haplotype estimations are further described below under the heading "Statistical methods". Any other method known in the art to determine or to estimate the frequency of a haplotype in a population may also be used.

ii. Linkage disequilibrium analysis

Linkage disequilibrium is the non-random association of alleles at two or more loci and represents a powerful tool for mapping genes involved in disease traits (see Ajioka R.S. et al., *Am. J. Hum. Genet.*, 60:1439-1447, 1997). Biallelic markers, because they are densely spaced in the human genome and can be genotyped in more numerous numbers than other types of genetic markers (such as RFLP or VNTR markers), are particularly useful in genetic analysis based on linkage disequilibrium. The biallelic markers of the present invention may be used in any linkage disequilibrium analysis method known in the art.

Briefly, when a disease mutation is first introduced into a population (by a new mutation or the immigration of a mutation carrier), it necessarily resides on a single chromosome and thus on a single "background" or "ancestral" haplotype of linked markers. Consequently, there is complete disequilibrium between these markers and the disease mutation: one finds the disease mutation only in the presence of a specific set of marker alleles. Through subsequent generations recombinations occur between the disease mutation and these marker polymorphisms, and the disequilibrium gradually dissipates. The pace of this dissipation is a function of the recombination frequency, so the markers closest to the disease gene will manifest higher levels of disequilibrium than those further away. When not broken up by recombination, "ancestral" haplotypes and linkage disequilibrium between marker alleles at different loci can be tracked not only through pedigrees but also through populations. Linkage disequilibrium is usually seen as an association between one specific allele at one locus and another specific allele at a second locus.

The pattern or curve of disequilibrium between disease and marker loci is expected to exhibit a maximum that occurs at the disease locus. Consequently, the amount of linkage disequilibrium between a disease allele and closely linked genetic markers may yield valuable information regarding the location of the disease gene. For fine-scale mapping of a disease locus, it is useful to have some knowledge of the patterns of linkage disequilibrium that exist between markers in the studied region. As mentioned above the mapping resolution achieved through the analysis of linkage disequilibrium is much higher than that of linkage studies. The

high density of biallelic markers combined with linkage disequilibrium analysis provides powerful tools for fine-scale mapping. Different methods to calculate linkage disequilibrium are described below under the heading "Statistical Methods".

iii. Population-based case-control studies of trait-marker associations

As mentioned above, the occurrence of pairs of specific alleles at different loci on the same chromosome is not random and the deviation from random is called linkage disequilibrium. Association studies focus on population frequencies and rely on the phenomenon of linkage disequilibrium. If a specific allele in a given gene is directly involved in causing a particular trait, its frequency will be statistically increased in an affected (trait positive) population, when compared to the frequency in a trait negative population or in a random control population. As a consequence of the existence of linkage disequilibrium, the frequency of all other alleles present in the haplotype carrying the trait-causing allele will also be increased in trait positive individuals compared to trait negative individuals or random controls. Therefore, association between the trait and any allele (specifically a biallelic marker allele) in linkage disequilibrium with the trait-causing allele will suffice to suggest the presence of a trait-related gene in that particular region. Case-control populations can be genotyped for biallelic markers to identify associations that narrowly locate a trait causing allele. As any marker in linkage disequilibrium with one given marker associated with a trait will be associated with the trait. Linkage disequilibrium allows the relative frequencies in case-control populations of a limited number of genetic polymorphisms (specifically biallelic markers) to be analyzed as an alternative to screening all possible functional polymorphisms in order to find trait-causing alleles. Association studies compare the frequency of marker alleles in unrelated case-control populations, and represent powerful tools for the dissection of complex traits.

1) Case-control populations (inclusion criteria)

Population-based association studies do not concern familial inheritance but compare the prevalence of a particular genetic marker, or a set of markers, in case-control populations. They are case-control studies based on comparison of unrelated case (affected or trait positive) individuals and unrelated control (unaffected or trait negative or random) individuals. Preferably the control group is composed of unaffected or trait negative individuals. Further, the control group is ethnically matched to the case population. Moreover, the control group is preferably matched to the case-population for the main known confusion factor for the trait under study (for example age-matched for an age-dependent trait). Ideally, individuals in the two samples are paired in such a way that they are expected to differ only in their disease status. In the following "trait positive population", "case population" and "affected population" are used interchangeably.

An important step in the dissection of complex traits using association studies is the choice of case-control populations (see Lander and Schork, *Science*, 265, 2037-2048, 1994). A

major step in the choice of case-control populations is the clinical definition of a given trait or phenotype. Any genetic trait may be analyzed by the association method proposed here by carefully selecting the individuals to be included in the trait positive and trait negative phenotypic groups. Four criteria are often useful: clinical phenotype, age at onset, family history and severity. The selection procedure for continuous or quantitative traits (such as blood pressure for example) involves selecting individuals at opposite ends of the phenotype distribution of the trait under study, so as to include in these trait positive and trait negative populations individuals with non-overlapping phenotypes. Preferably, case-control populations consist of phenotypically homogeneous populations. Trait positive and trait negative populations consist of phenotypically uniform populations of individuals representing each between 1 and 98%, preferably between 1 and 80%, more preferably between 1 and 50%, and more preferably between 1 and 30%, most preferably between 1 and 20% of the total population under study, and selected among individuals exhibiting non-overlapping phenotypes. The clearer the difference between the two trait phenotypes, the greater the probability of detecting an association with biallelic markers. The selection of those drastically different but relatively uniform phenotypes enables efficient comparisons in association studies and the possible detection of marked differences at the genetic level, provided that the sample sizes of the populations under study are significant enough.

In preferred embodiments, a first group of between 50 and 300 trait positive individuals, preferably about 100 individuals, are recruited according to their phenotypes. A similar number of trait negative individuals are included in such studies.

In the present invention, typical examples of inclusion criteria include a CNS disorder or the evaluation of the response to a drug acting on a CNS disorder or side effects to treatment with drugs acting on a CNS disorder.

Suitable examples of association studies using biallelic markers including the biallelic markers of the present invention, are studies involving the following populations:

a case population suffering from a CNS disorder and a healthy unaffected control population, or

a case population treated with agents acting on a CNS disorder suffering from side-effects resulting from the treatment and a control population treated with the same agents showing no side-effects, or

a case population treated with agents acting on a CNS disorder showing a beneficial response and a control population treated with same agents showing no beneficial response.

2) Association analysis

The general strategy to perform association studies using biallelic markers derived from a region carrying a candidate gene is to scan two groups of individuals (case-control populations)

in order to measure and statistically compare the allele frequencies of the biallelic markers of the present invention in both groups.

If a statistically significant association with a trait is identified for at least one or more of the analyzed biallelic markers, one can assume that: either the associated allele is directly responsible for causing the trait (the associated allele is the trait causing allele), or more likely the associated allele is in linkage disequilibrium with the trait causing allele. The specific characteristics of the associated allele with respect to the candidate gene function usually gives further insight into the relationship between the associated allele and the trait (causal or in linkage disequilibrium). If the evidence indicates that the associated allele within the candidate gene is most probably not the trait causing allele but is in linkage disequilibrium with the real trait causing allele, then the trait causing allele can be found by sequencing the vicinity of the associated marker.

Association studies are usually run in two successive steps. In a first phase, the frequencies of a reduced number of biallelic markers from one or several candidate genes are determined in the trait positive and trait negative populations. In a second phase of the analysis, the identity of the candidate gene and the position of the genetic loci responsible for the given trait is further refined using a higher density of markers from the relevant region. However, if the candidate gene under study is relatively small in length, as it is the case for many of the candidate genes analyzed included in the present invention, a single phase may be sufficient to establish significant associations.

3) Haplotype analysis

As described above, when a chromosome carrying a disease allele first appears in a population as a result of either mutation or migration, the mutant allele necessarily resides on a chromosome having a set of linked markers: the ancestral haplotype. This haplotype can be tracked through populations and its statistical association with a given trait can be analyzed. Complementing single point (allelic) association studies with multi-point association studies also called haplotype studies increases the statistical power of association studies. Thus, a haplotype association study allows one to define the frequency and the type of the ancestral carrier haplotype. A haplotype analysis is important in that it increases the statistical power of an analysis involving individual markers.

In a first stage of a haplotype frequency analysis, the frequency of the possible haplotypes based on various combinations of the identified biallelic markers of the invention is determined. The haplotype frequency is then compared for distinct populations of trait positive and control individuals. The number of trait positive individuals, which should be, subjected to this analysis to obtain statistically significant results usually ranges between 30 and 300, with a preferred number of individuals ranging between 50 and 150. The same considerations apply to the number of unaffected individuals (or random control) used in the study. The results of this

first analysis provide haplotype frequencies in case-control populations, for each evaluated haplotype frequency a p-value and an odd ratio are calculated. If a statistically significant association is found the relative risk for an individual carrying the given haplotype of being affected with the trait under study can be approximated.

4) Interaction analysis

The biallelic markers of the present invention may also be used to identify patterns of biallelic markers associated with detectable traits resulting from polygenic interactions. The analysis of genetic interaction between alleles at unlinked loci requires individual genotyping using the techniques described herein. The analysis of allelic interaction among a selected set of biallelic markers with appropriate level of statistical significance can be considered as a haplotype analysis. Interaction analysis consists in stratifying the case-control populations with respect to a given haplotype for the first loci and performing a haplotype analysis with the second loci with each subpopulation.

Statistical methods used in association studies are further described below in IV.C.

iv. Testing for linkage in the presence of association

The biallelic markers of the present invention may further be used in TDT (transmission/disequilibrium test). TDT tests for both linkage and association and is not affected by population stratification. TDT requires data from affected individuals and their parents or data from unaffected sibs instead of from parents (see Spielmann S. et al., *Am. J. Hum. Genet.*, 52:506-516, 1993; Schaid D.J. et al., *Genet. Epidemiol.*,13:423-450, 1996; Spielmann S. and Ewens W.J., *Am. J. Hum. Genet.*, 62:450-458, 1998). Such combined tests generally reduce the false – positive errors produced by separate analyses.

C. Statistical Methods

In general, any method known in the art to test whether a trait and a genotype show a statistically significant correlation may be used.

i. Methods in linkage analysis

Statistical methods and computer programs useful for linkage analysis are well-known to those skilled in the art (see Terwilliger J.D. and Ott J., *Handbook of Human Genetic Linkage, John Hopkins University Press, London*, 1994; Ott J., *Analysis of Human Genetic Linkage, John Hopkins University Press, Baltimore*, 1991).

ii. Methods to estimate haplotype frequencies in a population

As described above, when genotypes are scored, it is often not possible to distinguish heterozygotes so that haplotype frequencies cannot be easily inferred. When the gametic phase is not known, haplotype frequencies can be estimated from the multilocus genotypic data. Any method known to person skilled in the art can be used to estimate haplotype frequencies (see Lange K., *Mathematical and Statistical Methods for Genetic Analysis, Springer, New York*, 1997; Weir, B.S., *Genetic data Analysis II: Methods for Discrete population genetic Data, Sinauer*

Assoc., Inc., Sunderland, MA, USA, 1996). Preferably, maximum-likelihood haplotype frequencies are computed using an Expectation- Maximization (EM) algorithm (see Dempster et al., *J. R. Stat. Soc.*, 39B:1-38, 1977; Excoffier L. and Slatkin M., *Mol. Biol. Evol.*, 12(5): 921-927, 1995). This procedure is an iterative process aiming at obtaining maximum-likelihood estimates of haplotype frequencies from multi-locus genotype data when the gametic phase is unknown. Haplotype estimations are usually performed by applying the EM algorithm using for example the EM-HAPLO program (Hawley M.E. et al., *Am. J. Phys. Anthropol.*, 18:104, 1994) or the Arlequin program (Schneider et al., *Arlequin: a software for population genetics data analysis*, University of Geneva, 1997). The EM algorithm is a generalized iterative maximum likelihood approach to estimation and is briefly described below.

In what follows, phenotypes will refer to multi-locus genotypes with unknown haplotypic phase. Genotypes will refer to mutli-locus genotypes with known haplotypic phase.

Suppose one has a sample of N unrelated individuals typed for K markers. The data observed are the unknown-phase K-locus phenotypes that can be categorized with F different phenotypes. Further, suppose that we have H possible haplotypes (in the case of K biallelic markers, we have for the maximum number of possible haplotypes $H=2^K$).

For phenotype j with c_j possible genotypes, we have:

$$P_j = \sum_{i=1}^{c_j} P(genotype(i)) = \sum_{i=1}^{c_j} P(h_k, h_l). \qquad \text{Equation 1}$$

Here, P_j is the probability of the j^{th} phenotype, and $P(h_k,h_l)$ is the probability of the i^{th} genotype composed of haplotypes h_k and h_l. Under random mating (*i.e.* Hardy-Weinberg Equilibrium), $P(h_k h_l)$ is expressed as:

$$P(h_k, h_l) = P(h_k)^2 \text{ for } h_k = h_l, \text{ and}$$

$$P(h_k, h_l) = 2P(h_k)P(h_l) \text{ for } h_k \neq h_l. \qquad \text{Equation 2}$$

The E-M algorithm is composed of the following steps: First, the genotype frequencies are estimated from a set of initial values of haplotype frequencies. These haplotype frequencies are denoted $P_1^{(0)}, P_2^{(0)}, P_3^{(0)}, \ldots, P_H^{(0)}$. The initial values for the haplotype frequencies may be obtained from a random number generator or in some other way well known in the art. This step is referred to the Expectation step. The next step in the method, called the Maximization step, consists of using the estimates for the genotype frequencies to re-calculate the haplotype frequencies. The first iteration haplotype frequency estimates are denoted by $P_1^{(1)}, P_2^{(1)}, P_3^{(1)}, \ldots, P_H^{(1)}$. In general, the Expectation step at the s^{th} iteration consists of calculating the probability of placing each phenotype into the different possible genotypes based on the haplotype frequencies of the previous iteration:

$$P(h_k,h_l)^{(s)} = \frac{n_j}{N}\left[\frac{P_j(h_k,h_l)^{(s)}}{P_j}\right], \qquad \text{Equation 3}$$

where n_j is the number of individuals with the j^{th} phenotype and $P_j(h_k,h_l)^{(s)}$ is the probability of genotype $h_k h_l$ in phenotype j. In the Maximization step, which is equivalent to the gene-counting method (Smith, *Ann. Hum. Genet.*, 21:254-276, 1957), the haplotype frequencies are re-estimated based on the genotype estimates:

$$P_t^{(s+1)} = \frac{1}{2}\sum_{j=1}^{F}\sum_{i=1}^{c_j}\delta_{it}P_j(h_k,h_l)^{(s)}. \qquad \text{Equation 4}$$

Here, δ_{it} is an indicator variable which counts the number of occurrences that haplotype t is present in i^{th} genotype; it takes on values 0, 1, and 2.

The E-M iterations cease when the following criterion has been reached. Using Maximum Likelihood Estimation (MLE) theory, one assumes that the phenotypes j are distributed multinomially. At each iteration s, one can compute the likelihood function L. Convergence is achieved when the difference of the log-likehood between two consecutive iterations is less than some small number, preferably 10^{-7}.

iii. Methods to calculate linkage disequilibrium between markers

A number of methods can be used to calculate linkage disequilibrium between any two genetic positions, in practice linkage disequilibrium is measured by applying a statistical association test to haplotype data taken from a population.

Linkage disequilibrium between any pair of biallelic markers comprising at least one of the biallelic markers of the present invention (M_i, M_j) having alleles (a_i/b_i) at marker M_i and alleles (a_j/b_j) at marker M_j can be calculated for every allele combination (a_i,a_j; a_i,b_j; b_i,a_j and b_i,b_j), according to the Piazza formula :

$\Delta_{aiaj} = \sqrt{\theta 4} - \sqrt{(\theta 4 + \theta 3)(\theta 4 + \theta 2)}$, where :

$\theta 4$= - - = frequency of genotypes not having allele a_i at M_i and not having allele a_j at M_j

$\theta 3$= - + = frequency of genotypes not having allele a_i at M_i and having allele a_j at M_j

$\theta 2$= + - = frequency of genotypes having allele a_i at M_i and not having allele a_j at M_j

Linkage disequilibrium (LD) between pairs of biallelic markers (M_i, M_j) can also be calculated for every allele combination (ai,aj; ai,bj; b_i,a_j and b_i,b_j), according to the maximum-likelihood estimate (MLE) for delta (the composite genotypic disequilibrium coefficient), as described by Weir (Weir B.S., *Genetic Data Analysis, Sinauer Ass. Eds,* 1996). The MLE for the composite linkage disequilibrium is:

$D_{aiaj} = (2n_1 + n_2 + n_3 + n_4/2)/N - 2(pr(a_i).pr(a_j))$

Where $n_1 = \Sigma$ phenotype (a_i/a_i, a_j/a_j), $n_2 = \Sigma$ phenotype (a_i/a_i, a_j/b_j), $n_3 = \Sigma$ phenotype (a_i/b_i, a_j/a_j), n4= Σ phenotype (a_i/b_i, a_j/b_j) and N is the number of individuals in the sample.

This formula allows linkage disequilibrium between alleles to be estimated when only genotype, and not haplotype, data are available.

Another means of calculating the linkage disequilibrium between markers is as follows. For a couple of biallelic markers, M_i (a_i/b_i) and M_j (a_j/b_j), fitting the Hardy-Weinberg equilibrium, one can estimate the four possible haplotype frequencies in a given population according to the approach described above.

The estimation of gametic disequilibrium between *ai* and *aj* is simply:

$$D_{aiaj} = pr(haplotype(a_i, a_j)) - pr(a_i).pr(a_j).$$

Where $pr(a_i)$ is the probability of allele a_i and pr(a_j) is the probability of allele a_j and where $pr(haplotype\ (a_i, a_j))$ is estimated as in Equation 3 above.

For a couple of biallelic marker only one measure of disequilibrium is necessary to describe the association between M_i and M_j.

Then a normalised value of the above is calculated as follows:

$$D'_{aiaj} = D_{aiaj} / \max(-pr(a_i).pr(a_j), -pr(b_i).pr(b_j)) \quad \text{with } D_{aiaj}<0$$

$$D'_{aiaj} = D_{aiaj} / \max(pr(b_i).pr(a_j), pr(a_i).pr(b_j)) \quad \text{with } D_{aiaj}>0$$

The skilled person will readily appreciate that other LD calculation methods can be used without undue experimentation.

Linkage disequilibrium among a set of biallelic markers having an adequate heterozygosity rate can be determined by genotyping between 50 and 1000 unrelated individuals, preferably between 75 and 200, more preferably around 100.

iv. Testing for association

Methods for determining the statistical significance of a correlation between a phenotype and a genotype, in this case an allele at a biallelic marker or a haplotype made up of such alleles, may be determined by any statistical test known in the art and with any accepted threshold of statistical significance being required. The application of particular methods and thresholds of significance are well with in the skill of the ordinary practitioner of the art.

Testing for association is performed by determining the frequency of a biallelic marker allele in case and control populations and comparing these frequencies with a statistical test to determine if their is a statistically significant difference in frequency which would indicate a correlation between the trait and the biallelic marker allele under study. Similarly, a haplotype analysis is performed by estimating the frequencies of all possible haplotypes for a given set of biallelic markers in case and control populations, and comparing these frequencies with a statistical test to determine if their is a statistically significant correlation between the haplotype and the phenotype (trait) under study. Any statistical tool useful to test for a statistically significant association between a genotype and a phenotype may be used. Preferably the statistical test employed is a chi-square test with one degree of freedom. A p-value is then

determined (the P-value is the probability that a statistic as large or larger than the observed one would occur by chance).

1) Statistical significance

In preferred embodiments, significance for diagnostic purposes, either as a positive basis for further diagnostic tests or as a preliminary starting point for early preventive therapy, the p value related to a biallelic marker association is preferably about 1 x 10-2 or less, more preferably about 1 x 10-4 or less, for a single biallelic marker analysis and about 1 x 10-3 or less, still more preferably 1 x 10-6 or less and most preferably of about 1 x 10-8 or less, for a haplotype analysis involving several markers. These values are believed to be applicable to any association studies involving single or multiple marker combinations.

The skilled person can use the range of values set forth above as a starting point in order to carry out association studies with biallelic markers of the present invention. In doing so, significant associations between the biallelic markers of the present invention and CNS disorders can be revealed and used for diagnosis and drug screening purposes.

2) Phenotypic permutation

In order to confirm the statistical significance of the first stage haplotype analysis described above, it might be suitable to perform further analyses in which genotyping data from case-control individuals are pooled and randomized with respect to the trait phenotype. Each individual genotyping data is randomly allocated to two groups, which contain the same number of individuals as the case-control populations used to compile the data obtained in the first stage. A second stage haplotype analysis is preferably run on these artificial groups, preferably for the markers included in the haplotype of the first stage analysis showing the highest relative risk coefficient. This experiment is re-iterated preferably at least between 100 and 10000 times. The repeated iterations allow the determination of the percentage of obtained haplotypes with a significant p-value level.

3) Assessment of statistical association

To address the problem of false positives similar analysis may be performed with the same case-control populations in random genomic regions. Results in random regions and the candidate region are compared as described in US Provisional Patent Application entitled "Methods, software and apparati for identifying genomic regions harbouring a gene associated with a detectable trait".

v. Evaluation of risk factors

The association between a risk factor (in genetic epidemiology the risk factor is the presence or the absence of a certain allele or haplotype at marker loci) and a disease is measured by the odds ratio (OR) and by the relative risk (RR). If $P(R^+)$ is the probability of developing the disease for individuals with risk factor R and $P(R^-)$ is the probability for individuals without the risk factor, then the relative risk is simply the ratio of the two probabilities, that is:

$RR = P(R^+)/P(R^-)$

In case-control studies, direct measures of the relative risk cannot be obtained because of the sampling design. However, the odds ratio allows a good approximation of the relative risk for low-incidence diseases and can be calculated:

$$OR = \left[\frac{F^+}{1-F^+}\right] \Big/ \left[\frac{F^-}{(1-F^-)}\right]$$

$OR = [F^+/(1-F^+)] / [F^-/(1-F^-)]$

F^+ is the frequency of the exposure to the risk factor in cases and F^- is the frequency of the exposure to the risk factor in controls. F^+ and F^- are calculated using the allelic or haplotype frequencies of the study and further depend on the underlying genetic model (dominant, recessive, additive...).

One can further estimate the attributable risk (AR) which describes the proportion of individuals in a population exhibiting a trait due to a given risk factor. This measure is important in quantitating the role of a specific factor in disease etiology and in terms of the public health impact of a risk factor. The public health relevance of this measure lies in estimating the proportion of cases of disease in the population that could be prevented if the exposure of interest were absent. AR is determined as follows:

$AR = P_E (RR-1) / (P_E (RR-1)+1)$

AR is the risk attributable to a biallelic marker allele or a biallelic marker haplotype. P_E is the frequency of exposure to an allele or a haplotype within the population at large; and RR is the relative risk which is approximated with the odds ratio when the trait under study has a relatively low incidence in the general population.

D. Association of Biallelic Markers of the Invention with Major Depression

In the context of the present invention, an association between biallelic marker alleles from candidate genes of the present invention and a CNS disorder was demonstrated. The considered CNS disorder was major depression.

Depression is a serious medical illness that affects 340 million people worldwide. In contrast to the normal emotional experiences of sadness, loss, or passing mood states, clinical depression is persistent and can interfere significantly with an individual's ability to function. Many neurochemical findings are coming to light implicating a biological basis for the depression, at least for certain subtypes. Abnormalities of monoamine function as well as over stimulation of the HPA axis have been recognized in depression for many years. Patterns of clustering and segregation in depressive families have suggested a genetic component to depression. However, the lack of a defined and specific depression phenotype and of suitable markers for genetic analysis is proving to be a major hurdle for reliably identifying genes associated with depression. As a result, psychiatrists today have to choose antidepressant

medications by intuition and trial and error; a situation that can put suicidal patients in jeopardy for weeks or months until the right compound is selected. Clearly, there is a strong need to successfully identify genes involved in depression; thus allowing researchers to understand the etiology of depression and address its cause, rather than symptoms.

As mentioned above, both the nervous system and endocrine system play a major role in the etiology of depression. More specifically, the neurotransmitters dopamine, norepinephrine and serotonin as well as the hormones corticotrophin releasing factor, glucocorticoids, mineralocorticoids and various neuropeptides are thought to play a major role in the pathophysiology of depression.

In order to investigate and identify a genetic origin of depression, a candidate gene scan for depression was conducted. The rational of this approach was to: 1) select candidate genes potentially involved in the pathophysiology of interest, in this case major depression, 2) to identify biallelic markers in those genes and finally 3) to measure the frequency of biallelic marker alleles in order to determine if some alleles are more frequent in depressed populations than in non-affected populations. Results were further validated by haplotype studies. Significant associations between biallelic marker alleles from the serotonin receptor 6 (5HTR6), serotonin 7 (5HTR7), serotonin transporter (5HTT), dopamine receptor 3 (DRD3), norepinephrine transporter (NET), guanine nucleotide binding protein, β3 (Gbeta3), glucocorticoid receptor (GRL), drug metabolizing enzyme cytochrome P450 3A4 (CYP3A4) and Wolfram Syndrome 1 (WFS1) genes and depression were demonstrated in the context of the present invention. Association studies are further described in Examples 3, 4 and 5.

This information is extremely valuable. The knowledge of a potential genetic predisposition, even if this predisposition is not absolute, might contribute in a very significant manner to treatment efficacy of depressed patients and to the development of diagnostic tools.

E. Identification of Biallelic Markers in Linkage Disequilibrium with the Biallelic Markers of the Invention

Once a first biallelic marker has been identified in a genomic region of interest, the practitioner of ordinary skill in the art, using the teachings of the present invention, can easily identify additional biallelic markers in linkage disequilibrium with this first marker. As mentioned before any marker in linkage disequilibrium with a first marker associated with a trait will be associated with the trait. Therefore, once an association has been demonstrated between a given biallelic marker and a trait, the discovery of additional biallelic markers associated with this trait is of great interest in order to increase the density of biallelic markers in this particular region. The causal gene or mutation will be found in the vicinity of the marker or set of markers showing the highest correlation with the trait.

Identification of additional markers in linkage disequilibrium with a given marker involves: (a) amplifying a genomic fragment comprising a first biallelic marker from a plurality

of individuals; (b) identifying of second biallelic markers in the genomic region harboring said first biallelic marker; (c) conducting a linkage disequilibrium analysis between said first biallelic marker and second biallelic markers; and (d) selecting said second biallelic markers as being in linkage disequilibrium with said first marker. Subcombinations comprising steps (b) and (c) are also contemplated.

Methods to identify biallelic markers and to conduct linkage disequilibrium analysis are described herein and can be carried out by the skilled person without undue experimentation. The present invention then also concerns biallelic markers which are in linkage disequilibrium with the specific biallelic markers shown in Table 7 and which are expected to present similar characteristics in terms of their respective association with a given trait.

F. Identification of Functional Mutations

Once a positive association is confirmed with a biallelic marker of the present invention, the associated candidate gene can be scanned for mutations by comparing the sequences of a selected number of trait positive and trait negative individuals. In a preferred embodiment, functional regions such as exons and splice sites, promoters and other regulatory regions of the candidate gene are scanned for mutations. Preferably, trait positive individuals carry the haplotype shown to be associated with the trait and trait negative individuals do not carry the haplotype or allele associated with the trait. The mutation detection procedure is essentially similar to that used for biallelic site identification.

The method used to detect such mutations generally comprises the following steps: (a) amplification of a region of the candidate gene comprising a biallelic marker or a group of biallelic markers associated with the trait from DNA samples of trait positive patients and trait negative controls; (b) sequencing of the amplified region; (c) comparison of DNA sequences from trait-positive patients and trait-negative controls; and (d) determination of mutations specific to trait-positive patients. Subcombinations which comprise steps (b) and (c) are specifically contemplated.

It is preferred that candidate polymorphisms be then verified by screening a larger population of cases and controls by means of any genotyping procedure such as those described herein, preferably using a microsequencing technique in an individual test format. Polymorphisms are considered as candidate mutations when present in cases and controls at frequencies compatible with the expected association results.

VII. Biallelic Markers of the Invention in Methods of Genetic Diagnostics

The biallelic markers of the present invention can also be used to develop diagnostics tests capable of identifying individuals who express a detectable trait as the result of a specific genotype or individuals whose genotype places them at risk of developing a detectable trait at a subsequent time. The trait analyzed using the present diagnostics may be any detectable trait,

including a CNS disorder, a response to an agent acting on a CNS disorder or side effects to an agent acting on a CNS disorder.

The diagnostic techniques of the present invention may employ a variety of methodologies to determine whether a test subject has a biallelic marker pattern associated with an increased risk of developing a detectable trait or whether the individual suffers from a detectable trait as a result of a particular mutation, including methods which enable the analysis of individual chromosomes for haplotyping, such as family studies, single sperm DNA analysis or somatic hybrids.

The present invention provides diagnostic methods to determine whether an individual is at risk of developing a disease or suffers from a disease resulting from a mutation or a polymorphism in a candidate gene of the present invention. The present invention also provides methods to determine whether an individual is likely to respond positively to an agent acting on a CNS disorder or whether an individual is at risk of developing an adverse side effect to an agent acting on a CNS disorder.

These methods involve obtaining a nucleic acid sample from the individual and, determining, whether the nucleic acid sample contains at least one allele or at least one biallelic marker haplotype, indicative of a risk of developing the trait or indicative that the individual expresses the trait as a result of possessing a particular candidate gene polymorphism or mutation (trait-causing allele).

Preferably, in such diagnostic methods, a nucleic acid sample is obtained from the individual and this sample is genotyped using methods described herein. The diagnostics may be based on a single biallelic marker or on a group of biallelic markers.

In each of these methods, a nucleic acid sample is obtained from the test subject and the biallelic marker pattern of one or more of the biallelic markers listed in Table 7 is determined.

In one embodiment, PCR amplification is conducted on the nucleic acid sample to amplify regions in which polymorphisms associated with a detectable phenotype have been identified. The amplification products are sequenced to determine whether the individual possesses one or more polymorphisms associated with a detectable phenotype. The primers used to generate amplification products may comprise the primers listed in Table 13. Alternatively, the nucleic acid sample is subjected to microsequencing reactions as described above to determine whether the individual possesses one or more polymorphisms associated with a detectable phenotype resulting from a mutation or a polymorphism in a candidate gene. The primers used in the microsequencing reactions may include the primers listed in Table 12. In another embodiment, the nucleic acid sample is contacted with one or more allele specific oligonucleotide probes which, specifically hybridize to one or more candidate gene alleles associated with a detectable phenotype. The probes used in the hybridization assay may include the probes listed in Table 14.

In a preferred embodiment the identity of the nucleotide present at, at least one, 5HTR6 related biallelic marker selected from the group consisting of 99-27207-117, 99-28110-75, and 99-28134-215, is determined and the detectable trait is depression.

In a preferred embodiment the identity of the nucleotide present at, at least one, 5HTR7 related biallelic marker selected from the group consisting of 99-32181-192 and 99-28106-185, is determined and the detectable trait is depression.

In a preferred embodiment the identity of the nucleotide present at, at least one, GRL related biallelic marker selected from the group consisting of 99-30858-354, 18-20-174, 99-32002-313, 18-31-178, 18-38-395, and 99-30853-364, is determined and the detectable trait is depression.

In a preferred embodiment the identity of the nucleotide present at, at least one, NET related biallelic marker selected from the group consisting of 19-56-140, 19-28-136, 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, and 16-50-196, is determined and the detectable trait is depression.

In a preferred embodiment the identity of the nucleotide present at, at least one, DRD3 related biallelic marker selected from the group consisting of 8-19-372, is determined and the detectable trait is depression.

In a preferred embodiment the identity of the nucleotide present at, at least one, CYP3A4 related biallelic marker selected from the group consisting of 12-254-180, 10-214-279, and 10-217-91, is determined and the detectable trait is depression.

In a preferred embodiment the identity of the nucleotide present at, at least one, 5HTT related biallelic marker selected from the group consisting of 18-194-130, 18-186-391, 18-198-252, and 18-242-300, is determined and the detectable trait is depression.

In a preferred embodiment the identity of the nucleotide present at, at least one, Gbeta3 related biallelic marker selected from the group consisting of 20-205-302, 19-58-162, 19-9-45, 19-22-74, and 19-88-185, is determined and the detectable trait is depression.

In a preferred embodiment the identity of the nucleotide present at, at least one, WFS1 related biallelic marker selected from the group consisting of 19-18-310, 19-19-174, 19-17-188, and 19-16-127, is determined and the detectable trait is depression.

Diagnostic kits comprising polynucleotides of the present invention are further described in section I.

These diagnostic methods are extremely valuable as they can, in certain circumstances, be used to initiate preventive treatments or to allow an individual carrying a significant haplotype to foresee warning signs such as minor symptoms. In diseases in which attacks may be extremely violent and sometimes fatal if not treated on time, such as asthma, the knowledge of a potential predisposition, even if this predisposition is not absolute, might contribute in a very significant manner to treatment efficacy. Similarly, a diagnosed predisposition to a potential side

effect could immediately direct the physician toward a treatment for which such side effects have not been observed during clinical trials.

Diagnostics, which analyze and predict response to a drug or side effects to a drug, may be used to determine whether an individual should be treated with a particular drug. For example, if the diagnostic indicates a likelihood that an individual will respond positively to treatment with a particular drug, the drug may be administered to the individual. Conversely, if the diagnostic indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment may be prescribed. A negative response may be defined as either the absence of an efficacious response or the presence of toxic side effects.

Clinical drug trials represent another application for the markers of the present invention. One or more markers indicative of response to an agent acting on a CNS disorder or to side effects to an agent acting on a CNS disorder may be identified using the methods described above. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems.

VIII. DNA Typing Methods and Systems

The present invention also encompasses a DNA typing system having a much higher discriminatory power than currently available typing systems. The systems and associated methods are particularly applicable in the identification of individuals for forensic science and paternity determinations. These applications have become increasingly important; in forensic science, for example, the identification of individuals by polymorphism analysis has become widely accepted by courts as evidence.

While forensic geneticists have developed many techniques to compare homologous segments of DNA to determine if the segments are identical or if they differ in one or more nucleotides, each technique still has certain disadvantages. In particular, the techniques vary widely in terms of expense of analysis, time required to carry out an analysis and statistical power.

RFLP analysis methods

The best known and most widespread method in forensic DNA typing is the restriction fragment length polymorphism (RFLP) analysis. In RFLP testing, a repetitive DNA sequence referred to as a variable number tandem repeat (VNTR) which varies between individuals is analyzed. The core repeat is typically a sequence of about 15 base pairs in length, and highly polymorphic VNTR loci can have an average of about 20 alleles. DNA restriction sites located

on either site of the VNTR are exploited to create DNA fragments from about 0.5Kb to less than 10Kb which are then separated by electrophoresis, indicating the number of repeats found in the individual at the particular loci. RFLP methods generally consist of (1) extraction and isolation of DNA, (2) restriction endonuclease digestion; (3) separation of DNA fragments by electrophoresis; (4) capillary transfer; (5) hybridization with radiolabelled probes; (6) autoradiography; and (7) interpretation of results (Lee, H.C. et al., Am. J. Forensic. Med. Pathol. 15(4): 269-282 (1994)). RFLP methods generally combine analysis at about 5 loci and have much higher discriminate potential than other available test due the highly polymorphic nature of the VNTRs. However, autoradiography is costly and time consuming and an analysis generally takes weeks or months for turnaround. Additionally, a large amount of sample DNA is required, which is often not available at a crime scene. Furthermore, the reliability of the system and its credibility as evidence is decreased because the analysis of tightly spaced bands on electrophoresis results in a high rate of error.

PCR methods

PCR based methods offer an alternative to RFLP methods. In a first method called AmpFLP, DNA fragments containing VNTRs are amplified and then separated electrophoretically, without the restriction step of RFLP method. While this method allows small quantities of sample DNA to be used, decreases analysis time by avoiding autoradiography, and retains high discriminatory potential, it nevertheless requires electrophoretic separation which takes substantial time and introduces an significant error rate. In another AmpFLP method, short tandem repeats (STRs) of 2 to 8 base pairs are analyzed. STRs are more suitable to analysis of degraded DNA samples since they require smaller amplified fragments but have the disadvantage of requiring separation of the amplified fragments. While STRs are far less informative than longer repeats, similar discriminatory potential can be achieved if enough STRs are used in a single analysis.

Other methods include sequencing of mitochondrial DNA, which is especially suitable for situations where sample DNA is very degraded or in small quantities. However, only a small region of 1Kb of the mitochondrial DNA referred to as the D-Loop locus has been found useful for typing because of its polymorphic nature, resulting in lower discriminatory potential than with RFLP or AmpFLP methods. Furthermore, DNA sequencing is expensive to carry out on a large number of samples.

Further available methods include dot-blot methods, which involve using allele specific oligonucleotide probes which hybridize sequence specifically to one allele of a polymorphic site. Systems include the HLA DQ-alpha kit developed by Cetus Corp. which has a discriminatory value of about 1 in 20, and a dot-blot strip referred to as the Polymarker strip combining five genetic loci for a discriminatory value of about one in a few thousand. (Weedn, V., Clinics in Lab. Med. 16(1): 187-196 (1996)).

In addition to difficulties in analysis and time consuming laboratory procedures, it remains desirable for all DNA typing systems to have a higher discriminatory power. Several applications exist in which even the most discriminating tests need improvement in order to remove the considerable remaining doubt resulting from such analyses. Table 3 below lists characteristics of currently available forensic testing systems (Weedn, (1996)) and compares them with the method of the invention.

Table 3

Test type	*Technology*	*Turnaround time*	*Discriminatory potential*	*Sensitivity (amount DNA)*	*Sample*
RFLP	VNTR (autoradiography)	Weeks or months	10^6 to 10^9	10ng	Highly intact DNA
AmpFLP	VNTR (PCR based)	Days	10^3 to 10^6	100pg	Moderate degradation
Dot blot (ex. HLADQA1)	Sequence specific oligonucleotide probes	Days	10^1 to 10^3	1ng	Moderate degradation
Mitochondrial DNA	D-loop sequence (PCR based)	Days	10^2	1pg	Severe degradation
Present marker set	Biallelic Markers (set of 13, set of 100, set of 200, set of 270)	Hours to Days (throughput dependent)	10^6, 10^{47}, 10^{238}	100pg	Moderate degradation

Applications

As described above, an important application of DNA typing tests is to determine whether a DNA sample (e.g. from a crime scene) originated from an individual suspected of leaving said DNA sample.

There are several applications for DNA typing which require a particularly powerful genotyping system. In a first application, a high powered typing system is advantageous when for example a suspect is identified by searching a DNA profile database such as that maintained by the U.S. Federal Bureau of Investigation. Since databases may contain large numbers of data entries that are expected to increase consistently, currently used forensic systems can be expected to identify several matching DNA profiles due to their relative lack of power. While database searches generally reinforce the evidence by excluding other possible suspects, low powered typing systems resulting in the identification of several individuals may often tend to diminish the overall case against a defendant.

In another application, a target population is systematically tested to identify an individual having the same DNA profile as that of a DNA sample. In such a situation, a defendant is chosen at random based on DNA profile from a large population of innocent individuals. Since the population tested can often be large enough that at least one positive match is identified, and it is usually not possible to exhaustively test a population, the usefulness of the evidence will depend on the level of significance of the forensic test. In order to render such an application useful as a sole or primary source of evidence, DNA typing systems of extremely high discriminatory potential are required.

In yet another application, it is desirable to be able to discriminate between related individuals. Because related individuals will be expected to share a large portion of alleles at polymorphic sites, a very high powered DNA typing assay would be required to discriminate between them. This can have important effects if a sample is found to match the defendant's DNA profile and no evidence that the perpetrator is a relative can be found.

Accordingly, there a need in this art for a rapid, simple, inexpensive and accurate technique having a very high resolution value to determine relationships between individuals and differences in degree of relationships. Also, there is a need in the art for a very accurate genetic relationship test procedure which uses very small amounts of an original DNA sample, yet produces very accurate results.

The present invention thus involves methods for the identification of individuals comprising determining the identity of the nucleotides at set of genetic markers in a biological sample, wherein said set of genetic markers comprises at least one CNS disorder-related marker. The present invention provides an extensive set of biallelic markers allowing a higher discriminatory potential than the genetic markers used in current forensic typing systems. Also, biallelic markers can be genotyped in individuals with much higher efficiency and accuracy than the genetic markers used in current forensic typing systems. In preferred embodiments, the invention comprises determining the identity of a nucleotide at a CNS disorder-related marker by single nucleotide primer extension, which does not require electrophoresis as in techniques described above and results in lower rate of experimental error. As shown in Table 3, above, in comparison with PCR based VNTR based methods which allow discriminatory potential of thousands to millions, and RFLP based methods which allow discriminatory potential of merely millions to billions under optimal assumptions, the biallelic marker based method of the present invention provides a radical increase in discriminatory potential.

Any suitable set of genetic markers and biallelic markers of the invention may be used, and may be selected according to the discriminatory power desired. Biallelic markers, sets of biallelic markers, probes, primers, and methods for determining the identity of said biallelic markers are further described herein.

Discriminatory potential of biallelic marker typing

Calculating discriminatory potential

The discriminatory potential of the forensic test can be determined in terms of the profile frequency, also referred to as the random match probability, by applying the product rule. The product rule involves multiplying the allelic frequencies of all the individual alleles tested, and multiplying by an additional factor of 2 for each heterozygous locus.

In one example discussed below, the discriminatory potential of biallelic marker typing can be considered in the context of forensic science. In order to determine the discriminatory potential with respect to the numbers of biallelic markers to be used in a genetic typing system, the formulas and calculations below assume that (1) the population under study is sufficiently large (so that we can assume no consanguinity); (2) all markers chosen are not correlated, so that the product rule (Lander and Budlowle (1992)) can be applied; and (3) the ceiling rule can be applied or that the allelic frequencies of markers in the population under study are known with sufficient accuracy.

As noted in Weir, B.S., *Genetic data Analysis II: Methods for Discrete population genetic Data, Sinauer Assoc., Inc., Sunderland, MA, USA,* 1996, the example assumes a crime has been committed and a sample of DNA from the perpetrator (P) is available for analysis. The genotype of this DNA sample can be determined for several genetic markers, and the profile A of the perpetrator can thereby be determined.

In this example, one suspect (S) is available for typing. The same set of genetic markers, such as the biallelic markers of the invention, are typed and the same profile A is obtained for (S) and (P). Two hypotheses are thus presented as follows:

(1) either S is P (event C)

(2) either S is not P (event $\overline{C}$).

The ratio L of both probabilities can then be calculated using the following equation:

$$L = \frac{pr(S = A, P = A / C)}{pr(S = A, P = A / \overline{C})}$$

L can then further be calculated by the following equation:

$$L = \frac{1}{pr(P = A / S = A, \overline{C})} \quad (1)$$ Equation 1

These probabilities as well as L can be calculated in several settings, notably for different kinship coefficients between P and S for a genetic marker (see Weir, (1996)).

Assuming that all genetic markers chosen are independent of each other, the global ratio L for a set of genetic markers will be the product over each genetic marker of all L.

It is further possible to estimate the mean number of biallelic markers or VNTRs required to have a ratio L equal to 10^8 or 10^6 by calculating the expectancy of the random variable L using the following equation:

$E(L) = \prod_{i=1}^{N} E(L_i)$ where N is the number of loci

$E(L_i) = \sum_{j=1}^{G_i} pr(P = A_{ij} / S = A_{ij}, \overline{C}).L_{ij}$, where A_{ij} is the genotype j at the ith marker, L_{ij} the ratio associated with such genotype, G_i being the number of genotypes at locus i.

From equation 1, it can easily be derived that the expectancy of L_i is G_i, the number of possible genotypes of this marker.

The general expectancy for a set of genetic markers can then be expressed by the following equation:

$$E(L) = \prod_{i=1}^{N} G_i \quad (2)$$ Equation 2

Biallelic marker-based DNA typing systems

Using the equations described above, it is possible to select biallelic marker-based DNA typing systems having a desired discriminatory potential.

Using biallelic markers, E(L) can thus be expressed as 3^N. When using VNTR-based DNA typing systems, assuming the VNTRs have 10 alleles, E(L) can be expressed as 55^N. Based on these results, the number of biallelic markers or VNTRs needed to obtain, in mean, a ratio of at least 10^6 or 10^8 can calculated, and are set forth below in Table 4.

Table 4

Marker sets	L=10^6	L=10^8
Biallelic	13	17
5-allele markers (e.g. VNTR)	5	7
10-allele markers (e.g. VNTR)	4	5

Thus, in a first embodiment, DNA typing systems and methods of the invention may comprise genotyping a set of at least 13 or at least 17 biallelic markers to obtain a ratio of at least 10^6 or 10^8, assuming a flat distribution of L across the biallelic markers. In preferred embodiments, a greater number of biallelic markers is genotyped to obtain a higher L value. Preferably at least 1, 2, 3, 4, 5, 10, 13, 15, 17, 20, 25, 30, 40, 50, 70, 85, 100, 150, 200, 250 or all of the CNS disorder-related markers are genotyped. Said DNA typing systems of the invention would result in L values as listed in Table 5 below as an indication of the discriminate potential of the systems of the invention.

Table 5

Number of biallelic markers	L

50	$7.2 * 10^{23}$
100	$5*10^{47}$
271	3^271

In situations where the distribution of L is not flat, such as in the worst case when the perpetrator is homozygous for the major allele at each genetic locus and L thus takes the lowest value, a larger number of biallelic markers is required for the same discriminatory potential. Therefore, in preferred embodiments, DNA typing systems and methods of the invention using a larger number of biallelic markers allow for uneven distributions of L across the biallelic markers. For example, assuming unrelated individuals, a set of independent markers having an allelic frequency of 0.1/0.9, and the genetic profile of a homozygote at each genetic loci for the major allele, 66 biallelic markers are required to obtain a ratio of 10^6, and 88 biallelic markers are required to obtain a ratio of 10^8. Thus, in preferred embodiments based on the use of markers having a major allele of sufficiently high frequency, this is a first estimation of the upper bound of markers required in a DNA typing system.

In further embodiments, it is also desirable to have the ability to discriminate between relatives. Although unrelated individuals have a low probability of sharing genetic profiles, the probability is greatly increased for relatives. For example, the DNA profile of a suspect matches the DNA profile of a sample at a crime scene, and the probability of obtaining the same DNA profile if left by an untyped relative is required. Table 6 below (Weir (1996)) lists probabilities for several different types of relationships, assuming alleles A_i and A_j, and population frequencies p_i and p_j, and lists likelihood ratios assuming genetic loci having allele frequencies of 0.1.

Table 6

Genotype	Relationship	Pr(p=A|S=A)	L
$A_i A_j$	Full brothers	$(1+p_i+p_j+2p_i p_j)/4$	3.3
	Father and son	$(p_i+p_j)/2$	10.0
	Half brothers	$(p_i+p_j+4p_i p_j)/4$	16.7
	Uncle and nephew	$(1+p_i+p_j+2p_i p_j)/4$	16.7
	First cousins	$(1+p_i+p_j+12p_i p_j)/8$	25.0
	Unrelated	$2p_i p_j$	50.0
$A_j A_j$	Full brothers	$(1+p_i)^2/4$	3.3

Father and son	p_i	10.0
Half brothers	$p_i (1+p_i)/2$	18.2
Uncle and nephew	$p_i (1+p_i)/2$	18.2
First cousins	$p_i (1+3p_i)/4$	30.8
Unrelated	p_i^2	100.0

In one example, where the suspect is the full brother of the perpetrator, the number of required biallelic markers will be 187 assuming the profile is that of a homozygote for the major allele at each biallelic marker.

In yet further embodiments, the DNA typing systems and methods of the present invention may further take into account effects of subpopulations on the discriminatory potential. In embodiments described above for example, DNA typing systems consider close familial relationships, but do not take into account membership in the same population. While population membership is expected to have little effect, the invention may further comprise genotyping a larger set of biallelic markers to achieve higher discriminatory potential. Alternatively, a larger set of biallelic markers may be optimized for typing selected populations; alternatively, the ceiling principle may be used to study allele frequencies from individuals in various populations of interest, taking for any particular genotype the maximum allele frequency found among the populations.

The invention thus encompasses methods for genotyping comprising determining the identity of a nucleotide at least 13, 15, 17, 20, 25, 30, 40, 50, 66, 70, 85, 88, 100, 187, 200, or 250, 500, 700, 1000 or 2000 biallelic markers in a biological sample, wherein at least 1, 2, 3, 4, 5, 10, 13, 17, 20, 25, 30, 40, 50, 70, 85, 100, 150, 200, 250 or all of said biallelic markers are CNS disorder-related markers selected from the group consisting of SEQ ID NOS: 1-271.

Any markers known in the art may be used with the CNS disorder-related markers of the present invention in the DNA typing methods and systems described herein, for example in anyone of the following web sites offering collections of SNPs and information about those SNPs:

The Genetic Annotation Initiative (http://cgap.nci.nih.gov/GAI/). An NIH run site which contains information on candidate SNPs thought to be related to cancer and tumorigenesis generally.

dbSNP Polymorphism Repository (http://www.ncbi.nlm.nih.gov/SNP/). A more comprehensive NIH-run database containing information on SNPs with broad applicability in biomedical research.

HUGO Mutation Database Initiative (http://ariel.ucs.unimelb.edu.au:80/~cotton/mdi.htm). A database meant to provide systematic

access to information about human mutations including SNPs. This site is maintained by the Human Genome Organisation (HUGO).

Human SNP Database (http://www-genome.wi.mit.edu/SNP/human/index.html). Managed by the Whitehead Institute for Biomedical Research Genome Institute, this site contains information about SNPs resulting from the many Whitehead research projects on mapping and sequencing.

SNPs in the Human-Genome SNP database (http://www.ibc.wustl.edu/SNP). This website provides access to SNPs that have been organized by chromosomes and cytogenetic location. The site is run by Washington University.

HGBase (http://hgbase.cgr.ki.se/). HGBASE is an attempt to summarize all known sequence variations in the human genome, to facilitate research into how genotypes affect common diseases, drug responses, and other complex phenotypes, and is run by the Karolinska Institute of Sweden.

The SNP Consortium Database (http://snp.cshl.org/db/snp/map). A collection of SNPs and related information resulting from the collaborative effort of a number of large pharmaceutical and information processing companies.

GeneSNPs (http://www.genome.utah.edu/genesnps/). Run by the University of Utah, this site contains information about SNPs resulting from the U. S. National Institute of Environmental Health's initiative to understand the relationship between genetic variation and response to environmental stimuli and xenobiotics.

In addition, biallelic markers provided in the following patents and patent applications may also be used with the map-related biallelic markers of the invention in the DNA typing methods and systems described above: US Serial No. 60/206,615, filed 24 March 2000; US Serial No. 60/216,745, filed 30 June 2000; WIPO Serial No. PCT/IB00/00184, filed 11 February 2000; WIPO Serial No. PCT/IB98/01193, filed 17 July 1998; PCT Publication No. WO 99/54500, filed 21 April 1999; and WIPO Serial No. PCT/IB00/00403, filed 24 March 2000.

Biallelic markers, sets of biallelic markers, probes, primers, and methods for determining the identity of a nucleotide at said biallelic markers are also encompassed and are further described herein, and may encompass any further limitation described in this disclosure, alone or in any combination.

Forensic matching by microsequencing is further described in Example 6 below.

Throughout this application, various publications, patents, and published patent applications are cited. The disclosures of the publications, patents, and published patent specifications referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

EXAMPLES

Several of the methods of the present invention are described in the following examples, which are offered by way of illustration and not by way of limitation. Many other modifications and variations of the invention as herein set forth can be made without departing from the spirit and scope thereof and therefore only such limitations should be imposed as are indicated by the appended claims.

Example 1: *De Novo* Identification of Biallelic Markers

The biallelic markers set forth in this application were isolated from human genomic sequences. To identify biallelic markers, genomic fragments were amplified, sequenced and compared in a plurality of individuals.

DNA samples

Donors were unrelated and healthy. They represented a sufficient diversity for being representative of a French heterogeneous population. The DNA from 100 individuals was extracted and tested for the *de novo* identification of biallelic markers.

DNA samples were prepared peripheral venous blood as follows. 30 ml of peripheral venous blood were taken from each donor in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed in a lysis solution (50 ml final volume: 10 mM Tris pH7.6; 5 mM $MgCl_2$; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red cells present in the supernatant, after resuspension of the pellet in the lysis solution. The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of: (a) 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0.4 M; (b) 200 µl SDS 10%; and (c) 500 µl proteinase K (2 mg proteinase K in TE 10-2 / NaCl 0.4 M).

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm. For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the optical density (OD) at 260 nm (1 unit OD = 50 µg/ml DNA). To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent examples described below. DNA pools were constituted by mixing equivalent quantities of DNA from each individual.

Amplification of genomic DNA by PCR

Amplification of specific genomic sequences was carried out on pooled DNA samples obtained as described above.

Amplification primers

The primers used for the amplification of human genomic DNA fragments were defined with the OSP software (Hillier & Green, 1991). Preferably, primers included, upstream of the specific bases targeted for amplification, a common oligonucleotide tail useful for sequencing. Primers PU contain the following additional PU 5' sequence : TGTAAAACGACGGCCAGT; primers RP contain the following RP 5' sequence : CAGGAAACAGCTATGACC. Primers are listed in Table 12.

Amplification

PCR assays were performed using the following protocol:

Final volume		25 µl
DNA		2 ng/µl
$MgCl_2$		2 mM
dNTP (each)		200 µM
primer (each)		2.9 ng/µl
Ampli Taq Gold DNA polymerase		0.05 unit/µl
PCR buffer (10x = 0.1 M TrisHCl pH8.3 0.5M KCl)	1x	

DNA amplification was performed on a Genius II thermocycler. After heating at 94°C for 10 min, 40 cycles were performed. Cycling times and temperatures were: 30 sec at 94°C, 55°C for 1 min and 30 sec at 72°C. Holding for 7 min at 72°C allowed final elongation. The quantities of the amplification products obtained were determined on 96-well microtiter plates, using a fluorometer and Picogreen as intercalant agent (Molecular Probes).

Sequencing of amplified genomic DNA and identification of biallelic polymorphisms

Sequencing of the amplified DNA was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis (ABI Prism DNA Sequencing Analysis software 2.1.2 version).

The sequence data were further evaluated to detect the presence of biallelic markers within the amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position. However, the presence of two peaks can be an artifact due to background noise.

To exclude such an artifact, the two DNA strands were sequenced and a comparison between the two strands was carried out. In order to be registered as a polymorphic sequence, the polymorphism had to be detected on both strands. Further, some biallelic single nucleotide polymorphisms were confirmed by microsequencing as described below.

Biallelic markers were identified in the analyzed fragments and are shown in Table 7.

Example 2: Genotyping of Biallelic Markers

The biallelic markers identified as described above were further confirmed and their respective frequencies were determined through microsequencing. Microsequencing was carried out on individual DNA samples obtained as described herein.

Microsequencing primers

Amplification of genomic DNA fragments from individual DNA samples was performed as described in Example 1 using the same set of PCR primers (Table 12). Microsequencing was carried out on the amplified fragments using specific primers. See Table 13. The preferred primers used in microsequencing had about 19 nucleotides in length and hybridized just upstream of the considered polymorphic base.

The microsequencing reactions were performed as follows: 5 µl of PCR products were added to 5 µl purification mix (2U SAP (Shrimp alkaline phosphate) (Amersham E70092X)); 2U Exonuclease I (Amersham E70073Z); and 1 µl SAP buffer (200 mM Tris-HCl pH8, 100 mM $MgCl_2$) in a microtiter plate. The reaction mixture was incubated 30 minutes at 37°C, and denatured 10 minutes at 94°C afterwards. To each well was then added 20 µl of microsequencing reaction mixture containing: 10 pmol microsequencing oligonucleotide (19mers, GENSET, crude synthesis, 5 OD), 1 U Thermosequenase (Amersham E79000G), 1.25 µl Thermosequenase buffer (260 mM Tris HCl pH 9.5, 65 mM $MgCl_2$), and the two appropriate fluorescent ddNTPs complementary to the nucleotides at the polymorphic site corresponding to both polymorphic bases (11.25 nM TAMRA-ddTTP ; 16.25 nM ROX-ddCTP ; 1.675 nM REG-ddATP ; 1.25 nM RHO-ddGTP ; Perkin Elmer, Dye Terminator Set 401095). After 4 minutes at 94°C, 20 PCR cycles of 15 sec at 55°C, 5 sec at 72°C, and 10 sec at 94°C were carried out in a Tetrad PTC-225 thermocycler (MJ Research). The microtiter plate was centrifuged 10 sec at 1500 rpm. The unincorporated dye terminators were removed by precipitation with 19 µl $MgCl_2$ 2mM and 55 µl 100 % ethanol. After 15 minute incubation at room temperature, the microtiter plate was centrifuged at 3300 rpm 15 minutes at 4°C. After discarding the supernatants, the microplate was evaporated to dryness under reduced pressure (Speed Vac). Samples were resuspended in 2.5 µl formamide EDTA loading buffer and heated for 2 min at 95°C. 0.8 µl microsequencing reaction were loaded on a 10 % (19:1) polyacrylamide sequencing gel. The

data were collected by an ABI PRISM 377 DNA sequencer and processed using the GENESCAN software (Perkin Elmer).

Frequency of biallelic markers

Frequencies are reported for the less common allele only and are shown in Table 7.

Example 3: Association Study Between Major Depression and the Biallelic Markers of Candidate Genes

Collection of DNA samples from affected and non-affected individuals

The disease trait followed in this association study was major depression, a complex disorder believed to involve several neurotransmitter pathways including those utilizing norepinephrine and serotonin. The depressed patient population consists of 140 individuals that participated in a clinical study for the evaluation of the anti-depressant compound Reboxetine (Montgomery S.A. and Schatzberg A.F.; *Journal Clin. Psychiatry* 59(suppl 14): 3-7, 1998). Approximately 90% of these individuals were from a Caucasian ethnic background. The control population consisted of 94 individuals from a Caucasian population that had been found not to have any personal or family evidence of psychiatric disease.

Genotyping of affected and control individuals

The general strategy was to individually determine allele frequencies of biallelic markers in all individuals from each population described above. Allele frequencies of the biallelic markers were determined by performing microsequencing reactions on amplified DNA fragments obtained from genomic PCR performed on DNA samples from each individual. Genomic PCR and microsequencing were performed as detailed above in Examples 1 and 2.

Frequency of the biallelic markers alleles and genotypes of candidate gene and association with major depression

Frequencies of biallelic marker alleles were compared in the case-control populations described above. The data in Table 15 show the p-value obtained for each marker typed for each candidate gene for individual alleles and genotypes. Nine markers from 7 of 19 candidate genes were significant at the 5% level for allele frequency differences while seven markers from 6 of 19 candidate genes were significant at the 5% level for genotype frequency differences. In 4 cases, the same marker was significant at the 5% level for both allele and genotype frequency differences. This occurred for markers from the genes 5HTR6, 5HTR7, NET, and Gbeta3. These genes all participate in the mechanism of either serotonin or norepinephrine neurotransmission.

Haplotype frequency analysis

The results of the haplotype analysis using combinations of 2, 3, and 4 biallelic markers from each gene are shown in Tables 17 and 18. Haplotype analyses for the candidate genes were performed by estimating the frequencies of all 2, 3, and 4 marker haplotypes in the depressed and control populations. Haplotype estimations were performed by applying the Expectation-Maximization (EM) algorithm (Excoffier and Slatkin, *Mol. Biol. Evol.*, 12:921-927, 1995). Estimated haplotype frequencies in the depressed and control populations were compared by means of permutation tests based on individual haplotypes (Permutation test) as well as the distribution of frequencies from all possible haplotypes derived from a particular combination of given markers (Omnibus LR test).

The results of the Omnibus LR test are shown in Table 16. Listed are the top 10 marker combinations for each category of 4, 3, and 2 marker combinations and boxed in a double line border are the top 5% of each category (by p-value based on phenotypic reiteration of at least 1000 simulations). It is remarkable that several of the same genes identified by single marker association tests also appear in the top 5% of the Omnibus LR test. In particular, markers from the genes NET and Gbeta3 appear as top 5% in each category of combinations for the Omnibus LR test.

The results of the Permutation test for individual haplotypes are shown in Table 17. Listed are the top 20 haplotypes for each category of 4, 3, and 2 marker haplotypes and boxed in a double line border are the top 1% of each category (by p-value based on phenotypic reiteration of at least 1000 simulations). Again it is remarkable that several of the same genes identified by single marker association tests and Omnibus LR test also contribute haplotypes that appear in the top 1% of the Permutation test for individual haplotypes. Of all genes, only NET contributes to the top percentiles of each category of testing (individual markers for allele and genotype frequencies, Omnibus LR 4,3 and 2 marker combinations, and Permutation test for individual 4, 3, and 2 marker haplotypes). However several other genes contribute to several testing categories including previously mentioned Gbeta3 and 5HTR7 as well as WFS1, GRL, 5HTT and DRD3.

Two preferred haplotypes can be constructed from markers derived from the NET gene. One consists of markers 99-28788/300, 99-32061/304, and 99-32121/242 each manifesting the G allele. The GGG haplotype is present in only 1% of depressed cases vs. 7% of controls. While this haplotype is low in overall frequency, the p-value by permutation test is 2×10^{-4} and the p-value for this group of markers is 2×10^{-3} by Omnibus LR test suggesting that the result is highly significant. A second haplotype consists of markers 16-3/199, 16-28/93, and 16-50/196 manifesting alleles TCT respectively. The haplotype TCT is present in only 30% of cases vs. 43% of controls. The p-value by permutation test is 9×10^{-4} and the p-value for this group of markers is 8.9×10^{-4} by Omnibus LR test, also indicating a high level of significance.

Another example of a preferred haplotype comes from markers 16-16/285, 16-17/121, and 16-106/364 which are derived from the gene Gbeta3. The haplotype TTC is present in 21% of cases vs. 35% of controls. The p-value by permutation test is 1×10^{-3} and the p-value for this group of markers is 1×10^{-4} by Omnibus LR test, indicating a high level of significance.

Example 4: Association Study Between Major Depression and the Biallelic Markers of Candidate Genes

The association analysis of Example 3 was repeated using a different population set as described below. In general, these estimates agreed with the frequencies observed in the first screening within a few percent. Statistical assessments of haplotype frequency differences between depressed cases and controls were made by Omnibus LR tests and individual haplotype tests.

For Omnibus analyses, WFS1 marker combinations showed the most significant ($p<0.01$) differences between the depressed cases and controls for 2, 3, and 4 locus haplotypes. Strongest among these associations were combinations of markers spanning the core exonic region of the WFS1 gene including 19-17/188, 19-19/174, and 24-243/346. Several NET marker combinations showed significant associations ($p<0.05$) including those from the 5' flanking region and those from the exonic region. When compared to the distribution of Omnibus p-values observed in the 1st screening, 11 WFS1 marker combinations would have been among the top 5% of observed Omnibus p-values whereas 2 NET marker combinations would have been among the top 5%.

For individual haplotypes, haplotype GT from WFS1 markers 19-17/188 and 24-243/346 showed an 11% difference (37% cases vs. 26% controls, $p<0.001$). A similar difference was observed for haplotype GC from WFS1 markers 19-17/188 and 19-19/174 and GCT from all three markers ($p<0.005$). Several NET haplotypes showed >10% frequency differences between cases and controls ($p<0.01$). When compared to the distribution of individual haplotype p-values observed in the first screening, 6 WFS1 marker combinations would have been among the top 1% of observed individual haplotype p-values.

Frequency of the biallelic markers alleles and genotypes of candidate gene and association with major depression

Frequencies of biallelic marker alleles were compared in the case-control populations described above. The data in Table 18 show the p-value obtained for each marker typed for each candidate gene for individual alleles and genotypes. Nine markers from 7 of 19 candidate genes were significant at the 5% level for allele frequency differences while seven markers from 6 of 19 candidate genes were significant at the 5% level for genotype frequency differences. In 4 cases, the same marker was significant at the 5% level for both allele and genotype frequency differences. This occurred for markers from the genes 5HTR6, 5HTR7, and WFS1.

Haplotype frequency analysis

The results of the haplotype analysis using combinations of 2, 3, and 4 biallelic markers from each gene are shown in Tables 19 and 20. Haplotype analyses for the candidate genes were performed by estimating the frequencies of all 2, 3, and 4 marker haplotypes in the depressed and control populations. Haplotype estimations were performed by applying the Expectation-Maximization (EM) algorithm (Excoffier and Slatkin, *Mol. Biol. Evol.*, 12:921-927, 1995). Estimated haplotype frequencies in the depressed and control populations were compared by means of permutation tests based on individual haplotypes (Permutation test) as well as the distribution of frequencies from all possible haplotypes derived from a particular combination of given markers (Omnibus LR test).

The results of the Omnibus LR test are shown in Table 19. Listed are the top 10 marker combinations for each category of 4, 3, and 2 marker combinations and boxed in a double line border are the top 5% of each category (by p-value based on phenotypic reiteration of at least 1000 simulations). It is remarkable that several of the same genes identified by single marker association tests also appear in the top 5% of the Omnibus LR test. In particular, markers from the gene WFS1 appears as top 5% in each category of combinations for the Omnibus LR test.

The results of the Permutation test for individual haplotypes are shown in Table 20. Listed are the top 20 haplotypes for each category of 4, 3, and 2 marker haplotypes and boxed in a double line border are the top 1% of each category (by p-value based on phenotypic reiteration of at least 1000 simulations). Again it is remarkable that several of the same genes identified by single marker association tests and Omnibus LR test also contribute haplotypes that appear in the top 1% of the Permutation test for individual haplotypes. Of all genes, WFS1 contributes to the top percentiles of nearly all categories of testing (individual markers for allele and genotype frequencies, Omnibus LR 4,3 and 2 marker combinations, and Permutation test for individual 3 and 2 marker haplotypes). However several other genes contribute to several testing categories including previously mentioned 5HTR7 as well as NET, GRL, 5HTT and DRD3.

A preferred haplotype can be constructed from markers derived from the WFS1 gene. This consists of markers 19-17/188, 19-19/174, and 24-243/176 manifesting alleles GCT respectively. The GCT haplotype is present in 34% of depressed cases vs. 24% of controls. While this haplotype is low in overall frequency, the p-value by permutation test is 3×10^{-3} and the p-value for this group of markers is 1×10^{-3} by Omnibus LR test suggesting that the result is highly significant.

Example 5: Response to Reboxetine in Depressed Patients

Single point analyses were also performed on data from the candidate genes to determine Reboxetine response among depressed patients as compared to controls. Two markers from NET

(99-32061/304 and 99-32121/242) showed allelic and genotypic association ($p<0.05$). A single marker from Gbeta3 (18-355/67) showed allelic association with drug response ($p=0.05$).

Multipoint analyses on the data revealed Omnibus LR-based associations to be minimal for these genes with only two marker combinations from NET achieving a level of significance of $p<0.05$. At the individual haplotype level, a number of NET haplotypes achieved this level of significance with 10-15% responder/non-responder differences in haplotype frequencies. Also of note is the observation that a few individual haplotypes from Gbeta3 showed a remarkable level of significance ($p<0.0005$) corresponding to a nearly infinite relative risk (15-20% in non-responders vs. 0% in responders). This difference in estimated haplotype frequency is based largely on the observation that one particular haplotype cannot be unambiguously detected in the 204 responder haplotypes although there are at least 9 copies in 182 non-responder haplotypes.

In conclusion, modest association is present between NET and drug response. In addition, select individual haplotypes from Gbeta3 show a strong statistical association with drug response.

Example 6: Forensic Matching by Microsequencing

DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers based on a number of the sequences of SEQ ID NOS: 1 to 542 is then utilized according to the methods described herein to amplify DNA of approximately 500 bases in length from the forensic specimen. The alleles present at each of the selected biallelic markers site according to biallelic markers SEQ ID NOS: 1 to 542 are then identified according Example 2. A simple database comparison of the analysis results determines the differences, if any, between the sequences from a subject individual or from a database and those from the forensic sample. In a preferred method, statistically significant differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This lack of identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated with a large number of sequences, all matching. Preferably, a minimum of 13, 17, 20, 25, 30, 40, 50, 66, 70, 85, 88, 100, 187, 200 or 250 biallelic markers are used to test identity between the suspect and the sample.

In accordance with the regulations relating to Sequence Listings, the following codes have been used in the Sequence Listing to indicate the locations of biallelic markers within the sequences and to identify each of the alleles present at the polymorphic base. The code "r" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is an adenine. The code "y" in the sequences indicates that one allele of the polymorphic base is a thymine, while the other allele is a cytosine. The code "m" in the sequences indicates that one

allele of the polymorphic base is an adenine, while the other allele is an cytosine. The code "k" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a thymine. The code "s" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a cytosine. The code "w" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an thymine.

TABLE 7A

GENE	BIALLELIC MARKER ID	SEQ ID NO.	BIALLELIC MARKER POSITION IN SEQ ID NO.	VALIDATION MICRO-SEQUENCING	GENOTYPING LEAST COMMON ALLELE FREQUENCY	
5HTR6	99-27199-207	1	207	Y	T	0.30
5HTR6	99-27207-117	2	117	Y	C	0.37
5HTR6	99-27213-53	3	53	Y		
5HTR6	99-27218-333	4	333	Y		
5HTR6	99-28108-233	5	233	Y		
5HTR6	99-28109-275	6	275	N		
5HTR6	99-28110-75	7	74	Y	T	0.48
5HTR6	99-28125-81	8	81	Y		
5HTR6	99-28134-215	9	215	Y	T	0.37
5HTR6	99-28137-96	10	96	Y		
5HTR6	99-32204-305	11	305	N		
5HTR7	99-28149-118	12	118	Y	C	0.31
5HTR7	99-28160-285	13	285	Y	G	0.49
5HTR7	99-28171-458	14	457	Y	A	0.38
5HTR7	99-28173-395	15	395	Y		
5HTR7	99-32177-113	16	113	Y		
5HTR7	99-32181-192	17	192	Y	C	0.38
5HTR7	99-32193-258	18	257	Y	T	0.65
CHRNA7	99-28722-90	19	90	Y	C	0.32
CHRNA7	99-28730-351	20	351	Y	A	0.38
CHRNA7	99-32306-409	21	407	Y	G	0.33
CRFR1	99-27088-246	22	246	Y	A	0.38
CRFR1	99-27090-203	23	204	Y	G	0.22
CRFR1	99-27091-220	24	221	Y	A	0.49
CRFR1	99-27093-145	25	145	N		
CRFR1	99-27094-406	26	406	Y	T	0.21
CRFR1	99-27096-410	27	410	N		
CRFR1	99-27097-83	28	83	Y	T	0.47
CRFR1	99-27098-162	29	162	N		
CRFR1	99-27550-48	30	48	Y	A	0.26
CRFR1	99-27558-335	31	335	Y		
CRFR1	99-27561-106	32	106	Y		
CRFR1	99-27562-366	33	364	N		
MLR	16-31-738	34	738	Y	G	0.47
MLR	99-27110-301	35	300	Y		
MLR	99-27563-400	36	400	Y	A	0.44
MLR	99-27573-443	37	443	Y		
MLR	99-28732-133	38	133	Y	A	0.31
MLR	99-28735-56	39	56	Y	T	0.30
MLR	99-28736-399	40	399	Y		
MLR	99-28738-319	41	319	Y	C	0.37
MLR	99-28739-364	42	364	Y		
CRFR2	99-27875-185	43	185	Y	C	0.40
CRFR2	99-27880-176	44	176	Y	T	0.44
CRFR2	99-28747-371	45	373	Y	C	0.44
CRFR2	99-28753-353	46	352	Y	C	0.39
CRFR2	99-28755-206	47	207	Y	G	0.41
CRFR2	99-32333-366	48	366	N	C	
GRL	16-38-323	49	323	Y	A	0.33
GRL	99-28484-179	50	179	Y	A	0.40

TABLE 7A (cont)

GRL	99-30853-364	51	364	Y	G	0.42
GRL	99-28485-198	52	198	Y	G	0.20
GRL	99-30858-354	53	354	Y	T	0.16
GRL	99-32002-313	54	311	Y	A	0.48
GRL	18-15-366	55	366	Y		
GRL	18-20-174	56	174	Y	G	0.27
GRL	18-31-178	57	178	Y	C	0.35
GRL	18-38-395	58	395	Y	T	0.37
MAOA	18-2-192	59	192	Y	T	0.32
MAOB	99-26921-210	60	211	Y	G	0.48
MAOA	16-215-80	61	250	Y	T	0.33
MAOA-B	18-132-368	62	368	Y	C	0.34
MAOA	18-133-293	63	292	Y	A	0.27
5HTR2c	18-12-191	64	191	Y	A	0.14
5HTR2c	18-11-137	65	138	Y	G	0.27
5HTR2c	18-93-96	66	96	Y		
TH	16-115-343	67	343	Y	C	0.24
TH	16-42-140	68	140	Y	G	0.30
TH	18-251-176	69	176	Y	T	0.43
TH	18-269-44	70	44	Y	A	0.38
CRF	16-218-624	71	624	N		
CRF	18-393-330	72	330	Y		
CRF	18-394-402	73	402	Y		
DRD4	16-217-55	74	55	Y		
DRD4	18-284-139	75	139	Y		
DRD4	18-285-305	76	305	Y		
DRD4	18-289-239	77	239	Y		
DRD4	18-291-91	78	91	Y		
5HTT	18-186-391	79	391	Y	T	0.47
5HTT	18-194-130	80	130	Y	T	0.48
5HTT	18-198-252	81	252	Y	A	0.49
5HTT	18-242-300	82	299	Y	G	0.46
DRD3	8-15-126	83	1501	Y	G	0.30
DRD3	8-19-372	84	1501	Y	A	0.28
DRD3	99-2409-298	85	428	Y	A	0.40
DRD3	99-339-54	86	1501	Y	G	0.46
CYP3A4-7	12-254-180	87	311	Y	G	0.46
CYP3A4-7	10-214-279	88	1501	Y	C	0.12
CYP3A4-7	10-217-91	89	1501	Y	T	0.07
NET	99-28779-168	90	168	Y		
NET	99-28788-300	91	300	Y	A	0.47
NET	99-32052-262	92	263	Y		
NET	99-32121-242	93	244	Y	G	0.48
NET	99-32059-169	94	169	N		
NET	99-32061-304	95	304	Y	A	0.39
NET	99-32065-303	96	303	Y		
NET	99-32123-118	97	118	Y		
NET	99-32148-315	98	314	Y		
NET	16-2-76	99	95	Y	A	0.27
NET	16-28-93	100	120	Y	C	0.44
NET	16-3-199	101	342	Y	C	0.32
NET	16-50-197	102	197	Y	C	0.21
NET	16-1-59	103	181	Y		
NET	16-2-187	104	206	Y		
TACR1	99-28761-311	105	311	Y	A	0.22

TABLE 7A (ont)

TACR1	99-28771-86	106	86	Y	T	0.48
TACR1	99-28791-291	107	291	Y	A	0.26
TACR1	99-32077-66	108	66	Y		
TACR1	99-32078-466	109	467	Y		
TACR1	99-32376-426	110	426	Y		
TACR1	99-32361-419	111	420	Y	T	0.48
DRD2	16-21-228	112	228	Y	A	0.16
DRD2	16-22-156	113	156	Y	C	0.45
DRD2	16-23-404	114	404	Y	G	0.47
DRD2	16-24-175	115	175	Y	A	0.16
DRD2	16-25-286	116	286	Y	T	0.37
DRD2	16-25-279	117	279	Y		
DRD2	16-23-393	118	393	Y		
Gbeta3	16-106-364	119	364	Y	T	0.01
Gbeta3	16-16-285	120	285	Y	T	0.38
Gbeta3	16-17-121	121	121	Y	T	0.36
Gbeta3	16-84-185	122	185	Y	C	0.40
Gbeta3	16-87-74	123	74	Y	A	0.34
Gbeta3	16-91-333	124	333	Y	A	0.43
WFS1	16-128-142	125	142	Y	C	0.27
WFS1	16-133-205	126	245	Y	G	0.34
WFS1	16-135-181	127	232	Y	A	0.28
WFS1	16-145-405	128	455	Y	C	0.11
WFS1	16-177-320	129	320	Y	A	0.07
WFS1	16-4-354	130	354	Y	C	0.36

TABLE 7B

GENE	BIALLELIC MARKER ID	SEQ ID NO.	BIALLELIC MARKER POSITION IN SEQ ID NO.	VALIDATION MICRO-SEQUENCING	GENOTYPING LEAST COMMON ALLELE FREQUENCY	
5HTR6	99-27199-207	131	24	Y	T	0.30
5HTR6	99-27207-117	132	24	Y	C	0.37
5HTR6	99-27213-53	133	24	Y		
5HTR6	99-27218-333	134	24	Y		
5HTR6	99-28108-233	135	24	Y		
5HTR6	99-28109-275	136	24	N		
5HTR6	99-28110-75	137	24	Y	T	0.48
5HTR6	99-28125-81	138	24	Y		
5HTR6	99-28134-215	139	24	Y	T	0.37
5HTR6	99-28137-96	140	24	Y		
5HTR6	99-32204-305	141	24	N		
5HTR7	99-28149-118	142	24	Y	C	0.31
5HTR7	99-28160-285	143	24	Y	G	0.49
5HTR7	99-28171-458	144	24	Y	A	0.38
5HTR7	99-28173-395	145	24	Y		
5HTR7	99-32177-113	146	24	Y		
5HTR7	99-32181-192	147	24	Y	C	0.38
5HTR7	99-32193-258	148	24	Y	T	0.65
CHRNA7	99-28722-90	149	24	Y	C	0.32
CHRNA7	99-28730-351	150	24	Y	A	0.38
CHRNA7	99-32306-409	151	24	Y	G	0.33
CRFR1	99-27088-246	152	24	Y	A	0.38
CRFR1	99-27090-203	153	24	Y	G	0.22

TABLE 7B (c nt)

CRFR1	99-27091-220	154	24	Y	A	0.49
CRFR1	99-27093-145	155	24	N		
CRFR1	99-27094-406	156	24	Y	T	0.21
CRFR1	99-27096-410	157	24	N		
CRFR1	99-27097-83	158	24	Y	T	0.47
CRFR1	99-27098-162	159	24	N		
CRFR1	99-27550-48	160	24	Y	A	0.26
CRFR1	99-27558-335	161	24	Y		
CRFR1	99-27561-106	162	24	Y		
CRFR1	99-27562-366	163	24	N		
MLR	16-31-738	164	24	Y	G	0.47
MLR	99-27110-301	165	24	Y		
MLR	99-27563-400	166	24	Y	A	0.44
MLR	99-27573-443	167	24	Y		
MLR	99-28732-133	168	24	Y	A	0.31
MLR	99-28735-56	169	24	Y	T	0.30
MLR	99-28736-399	170	24	Y		
MLR	99-28738-319	171	24	Y	C	0.37
MLR	99-28739-364	172	24	Y		
CRFR2	99-27875-185	173	24	Y	C	0.40
CRFR2	99-27880-176	174	24	Y	T	0.44
CRFR2	99-28747-371	175	24	Y	C	0.44
CRFR2	99-28753-353	176	24	Y	C	0.39
CRFR2	99-28755-206	177	24	Y	G	0.41
CRFR2	99-32333-366	178	24	N	C	
GRL	16-38-323	179	24	Y	A	0.33
GRL	99-28484-179	180	24	Y	A	0.40
GRL	99-30853-364	181	24	Y	G	0.42
GRL	99-28485-198	182	24	Y	G	0.20
GRL	99-30858-354	183	24	Y	T	0.16
GRL	99-32002-313	184	24	Y	A	0.48
GRL	18-15-366	185	24	Y		
GRL	18-20-174	186	24	Y	G	0.27
GRL	18-31-178	187	24	Y	C	0.35
GRL	18-38-395	188	24	Y	T	0.37
MAOA	18-2-192	189	24	Y	T	0.32
MAOB	99-26921-210	190	24	Y	G	0.48
MAOA	16-215-80	191	24	Y	T	0.33
MAOA-B	18-132-368	192	24	Y	C	0.34
MAOA	18-133-293	193	24	Y	A	0.27
5HTR2c	18-12-191	194	24	Y	A	0.14
5HTR2c	18-11-137	195	24	Y	G	0.27
5HTR2c	18-93-96	196	24	Y		
TH	16-115-343	197	24	Y	C	0.24
TH	16-42-140	198	24	Y	G	0.30
TH	18-251-176	199	24	Y	T	0.43
TH	18-269-44	200	24	Y	A	0.38
CRF	16-218-624	201	24	N		
CRF	18-393-330	202	24	Y		
CRF	18-394-402	203	24	Y		
DRD4	16-217-55	204	24	Y		
DRD4	18-284-139	205	24	Y		
DRD4	18-285-305	206	24	Y		
DRD4	18-289-239	207	24	Y		
DRD4	18-291-91	208	24	Y		

TABLE 7B (c nt)

5HTT	18-186-391	209	24	Y	T	0.47
5HTT	18-194-130	210	24	Y	T	0.48
5HTT	18-198-252	211	24	Y	A	0.49
5HTT	18-242-300	212	24	Y	G	0.46
DRD3	8-15-126	213	24	Y	G	0.30
DRD3	8-19-372	214	24	Y	A	0.28
DRD3	99-2409-298	215	24	Y	A	0.40
DRD3	99-339-54	216	24	Y	G	0.46
CYP3A4-7	12-254-180	217	24	Y	G	0.46
CYP3A4-7	10-214-279	218	24	Y	C	0.12
CYP3A4-7	10-217-91	219	24	Y	T	0.07
NET	99-28779-168	220	24	Y		
NET	99-28788-300	221	24	Y	A	0.47
NET	99-32052-262	222	24	Y		
NET	99-32121-242	223	24	Y	G	0.48
NET	99-32059-169	224	24	N		
NET	99-32061-304	225	24	Y	A	0.39
NET	99-32065-303	226	24	Y		
NET	99-32123-118	227	24	Y		
NET	99-32148-315	228	24	Y		
NET	16-2-76	229	24	Y	A	0.27
NET	16-28-93	230	24	Y	C	0.44
NET	16-3-199	231	24	Y	C	0.32
NET	16-50-197	232	24	Y	C	0.21
NET	16-1-59	233	24	Y		
NET	16-2-187	234	24	Y		
TACR1	99-28761-311	235	24	Y	A	0.22
TACR1	99-28771-86	236	24	Y	T	0.48
TACR1	99-28791-291	237	24	Y	A	0.26
TACR1	99-32077-66	238	24	Y		
TACR1	99-32078-466	239	24	Y		
TACR1	99-32376-426	240	24	Y		
TACR1	99-32361-419	241	24	Y	T	0.48
DRD2	16-21-228	242	24	Y	A	0.16
DRD2	16-22-156	243	24	Y	C	0.45
DRD2	16-23-404	244	24	Y	G	0.47
DRD2	16-24-175	245	24	Y	A	0.16
DRD2	16-25-286	246	24	Y	T	0.37
DRD2	16-25-279	247	24	Y		
DRD2	16-23-393	248	24	Y		
Gbeta3	16-106-364	249	24	Y	T	0.01
Gbeta3	16-16-285	250	24	Y	T	0.38
Gbeta3	16-17-121	251	24	Y	T	0.36
Gbeta3	16-84-185	252	24	Y	C	0.40
Gbeta3	16-87-74	253	24	Y	A	0.34
Gbeta3	16-91-333	254	24	Y	A	0.43
WFS1	16-128-142	255	24	Y	C	0.27
WFS1	16-133-205	256	24	Y	G	0.34
WFS1	16-135-181	257	24	Y	A	0.28
WFS1	16-145-405	258	24	Y	C	0.11
WFS1	16-177-320	259	24	Y	A	0.07
WFS1	16-4-354	260	24	Y	C	0.36

TABLE 7C

GENE	BIALLELIC MARKER ID	SEQ ID NO.	BIALLELIC MARKER POSITION IN SEQ ID NO.	VALIDATION MICRO-SEQUENCING	GENOTYPING LEAST COMMON ALLELE FREQUENCY	
MAO A/B	18-473-362	261	362	Y	C	0.43
MAO A/B	99-12361-88	262	88	Y	C	0.36
MAO A/B	99-12368-335	263	335	Y	C	0.36
MAO A/B	99-12370-67	264	67	Y	A	0.29
NET	99-32148-315	265	314	Y	C	0.27
NET	19-46-322	266	322	Y	C	0.31
NET	19-47-315	267	315	Y	T	0.14
NET	19-51-347	268	346	Y		
NET	99-32052-262	269	263	Y	T	0.38
CYP3A4/7	10-213-292	270	1501	Y	G	0.11
5HTT	18-419-135	271	135	Y		
5HTT	18-424-419	272	419	Y		
5HTT	18-429-289	273	290	Y		
5HTT	18-246-256	274	256	Y	C	0.48
Gbeta3	18-355-67	275	68	Y	C	0.49
Gbeta3	18-353-267	276	266	Y	T	0.27
Gbeta3	18-338-305	277	306	Y	G	0.3
WFS1	24-243-346	278	1501	Y	T	0.3
WFS1	99-62531-351	279	1501	Y	T	0.36
WFS1	99-54279-152	280	1501	Y	G	0.44
DRD2	18-168-245	281	245	Y	A	0.45
DRD2	18-171-291	282	291	Y	C	0.37
DRD2	18-172-346	283	346	Y	T	0.45
DRD2	18-177-406	284	406	Y	T	0.37
HM74	18-298-338	285	338	Y	G	0.49
HM74	18-298-110	286	110	Y		
HM74	18-299-105	287	104	Y		
HM74	18-884-30	288	31	Y	C	0.26
HM74	18-299-343	289	342	Y		
HM74	99-61513-139	290	140	Y	A	0.26
HM74	99-61514-179	291	179	Y	G	0.27
HM74	99-61516-323	292	323	Y	C	0.33
CRHBP	18-204-70	293	70	Y	C	0.20
CRHBP	18-207-441	294	442	Y	C	0.41
CRHBP	18-210-65	295	65	Y		
CRHBP	18-212-200	296	200	Y	T	0.31
CRHBP	18-229-334	297	334	Y	T	0.33
CRHBP	18-230-332	298	332	Y	T	0.32
AVPR1A	18-966-378	299	378	Y	C	0.41
AVPR1A	18-987-308	300	307	Y	A	0.35
AVPR1A	18-1169-118	301	118	Y		
AVPR1A	18-1172-138	302	138	Y	G	0.16
AVPR1A	18-1173-92	303	92	Y	T	0.32
AVPR1A	18-1174-387	304	387	Y	C	0.21
AVPR1A	18-1175-416	305	416	Y	G	0.21
AVPR1A	18-542-146	306	146	Y	G	0.21
5HT1A	8-42-211	307	1501	Y	G	0.46
5HT1A	8-45-389	308	1501	Y	G	0.01
5HT1A	18-994-270	309	270	Y	C	0.25
5HT1A	18-912-165	310	165	Y	T	0.46
5HT1A	18-991-124	311	124	Y	T	0.45

TABLE 7C (ont)

5HT1A	18-920-219	312	219	Y		
5HT1A	18-911-312	313	312	Y		
5HT1A	99-65963-368	314	368	Y		
5HT1A	99-65966-225	315	225	Y		
5HT1A	99-65968-75	316	75	Y		
5HT1A	99-5069-331	317	1501	Y		
5HT1A	99-5070-176	318	175	Y	T	0.02
GABRG2	18-511-348	319	348	Y		
GABRG2	18-523-352	320	352	Y	C	0.49
GABRG2	18-545-478	321	480	Y	G	0.45
GABRG2	18-522-194	322	194	Y	G	0.32
GABRG2	18-524-284	323	284	Y	C	0.36
ADRB1R	18-626-52	324	52	Y	G	0.36
ADRB1R	18-629-189	325	189	Y	T	0.40
ADRB1R	18-1131-71	326	71	Y	T	0.41
ADRB1R	18-534-126	327	126	Y		
ADRB1R	18-596-59	328	59	Y		
ADRB1R	18-597-27	329	27	Y	A	0.20
GABRA5	18-730-203	330	203	Y	G	0.48
GABRA5	18-734-89	331	89	Y	C	0.48
GABRA5	18-895-321	332	321	Y	A	0.27
GABRA5	18-896-69	333	69	Y		
GABRA5	18-903-58	334	58	Y		
GOLF	18-590-216	335	216	Y	G	0.42
GOLF	18-817-436	336	433	Y	T	0.41
GOLF	18-829-85	337	85	Y	G	0.39
GOLF	18-832-387	338	387	Y		
GOLF	18-833-259	339	259	Y		
GOLF	18-839-271	340	271	Y	C	0.40
GOLF	18-770-194	341	194	Y	T	0.37
GOLF	18-771-302	342	302	Y	G	0.30
GOLF	18-827-53	343	53	Y	G	0.34
GOLF	18-768-318	344	318	Y	G	0.31
GOLF	18-769-26	345	26	Y	A	0.17
SLC6A3	18-709-321	346	320	Y	C	0.46
SLC6A3	18-714-280	347	281	Y	T	0.17
SLC6A3	18-843-271	348	271	Y	C	0.37
SLC6A3	18-850-265	349	265	Y	T	0.34
SLC6A3	18-853-296	350	296	Y	T	0.23
SLC6A3	18-867-331	351	332	Y	C	0.48
SLC6A3	18-877-73	352	73	Y		
SLC6A3	18-856-85	353	85	Y	C	0.42
SLC6A3	18-861-101	354	101	Y	T	0.33
PDE4b	18-635-323	355	323	Y		
PDE4b	18-636-205	356	205	Y	A	0.36
PDE4b	18-649-427	357	427	Y	C	0.46
PDE4b	18-1134-316	358	316	Y	G	0.38
PDE4b	18-633-316	359	316	Y		
COMT	18-489-425	360	425	Y		
COMT	18-492-212	361	212	Y	C	0.41
COMT	18-488-156	362	156	Y	T	0.45
COMT	18-491-266	363	266	Y	T	0.42
COMT	18-497-141	364	141	Y	T	0.43
COMT	18-503-174	365	174	Y	C	0.31
COMT	18-490-95	366	95	Y	A	0.31

TABLE 7C (c nt)

NPY1R	18-699-115	367	114	Y		
NPY1R	18-1099-293	368	293	Y		
NPY1R	18-1105-22	369	22	Y		
SLC1	18-562-418	370	418	Y	C	0.49
SLC1	18-564-204	371	204	Y	C	0.50
SEF2-1B	18-1032-262	372	261	Y	C	0.33
SEF2-1B	18-1035-412	373	412	Y	C	0.50
SEF2-1B	18-1036-293	374	293	Y	C	0.34
SEF2-1B	18-1038-95	375	95	Y	T	0.34
SEF2-1B	18-1040-361	376	361	Y	G	0.44
SEF2-1B	18-748-356	377	356	Y	T	0.47
BDNF	18-937-181	378	179	Y	A	0.26
BDNF	18-942-175	379	175	Y	T	0.29
BDNF	18-1213-221	380	221	Y		
BDNF	18-937-147	381	145	Y		
BDNF	18-946-408	382	407	Y	C	0.20
GAP43	18-787-133	383	133	Y	A	0.39
GAP43	18-1149-239	384	239	Y	A	0.49
GAP43	18-1159-291	385	291	Y	G	0.23
GAP43	18-1135-273	386	273	Y	T	0.43
GAP43	18-1136-108	387	108	Y		
GAP43	18-1147-68	388	68	Y		
GAP43	18-1157-295	389	295	Y		
GAP43	18-802-460	390	459	Y	A	0.32
CLOCK	18-1064-110	391	109	Y	C	0.36
CLOCK	18-1068-327	392	327	Y	T	0.32
CLOCK	18-1069-365	393	365	Y	A	0.23
CLOCK	18-1073-367	394	367	Y	A	0.35
CLOCK	18-1070-272	395	272	Y	T	0.36
CLOCK	18-1057-35	396	35	Y		
CLOCK	18-1062-415	397	415	Y		
CLOCK	18-1082-165	398	165	Y		
CLOCK	18-1080-361	399	363	Y	C	0.18
HSP70	18-506-297	400	297	Y		
HSP70	18-570-38	401	38	Y		

TABLE 7D

GENE	BIALLELIC MARKER ID	SEQ ID NO.	BIALLELIC MARKER POSITION IN SEQ ID NO.	VALIDATION MICRO-SEQUENCING	GENOTYPING LEAST COMMON ALLELE FREQUENCY	
MAO A/B	18-473-362	402	24	Y	C	0.43
MAO A/B	99-12361-88	403	24	Y	C	0.36
MAO A/B	99-12368-335	404	24	Y	C	0.36
MAO A/B	99-12370-67	405	24	Y	A	0.29
NET	99-32148-315	406	24	Y	C	0.27
NET	19-46-322	407	24	Y	C	0.31
NET	19-47-315	408	24	Y	T	0.14
NET	19-51-347	409	24	Y		
NET	99-32052-262	410	24	Y	T	0.38
CYP3A4/7	10-213-292	411	24	Y	G	0.11
5HTT	18-419-135	412	24	Y		
5HTT	18-424-419	413	24	Y		
5HTT	18-429-289	414	24	Y		
5HTT	18-246-256	415	24	Y	C	0.48

TABLE 7D (cont)

Gbeta3	18-355-67	416	24	Y	C	0.49
Gbeta3	18-353-267	417	24	Y	T	0.27
Gbeta3	18-338-305	418	24	Y	G	0.3
WFS1	24-243-346	419	24	Y	T	0.3
WFS1	99-62531-351	420	24	Y	T	0.36
WFS1	99-54279-152	421	24	Y	G	0.44
DRD2	18-168-245	422	24	Y	A	0.45
DRD2	18-171-291	423	24	Y	C	0.37
DRD2	18-172-346	424	24	Y	T	0.45
DRD2	18-177-406	425	24	Y	T	0.37
HM74	18-298-338	426	24	Y	G	0.49
HM74	18-298-110	427	24	Y		
HM74	18-299-105	428	24	Y		
HM74	18-884-30	429	24	Y	C	0.26
HM74	18-299-343	430	24	Y		
HM74	99-61513-139	431	24	Y	A	0.26
HM74	99-61514-179	432	24	Y	G	0.27
HM74	99-61516-323	433	24	Y	C	0.33
CRHBP	18-204-70	434	24	Y	C	0.20
CRHBP	18-207-441	435	24	Y	C	0.41
CRHBP	18-210-65	436	24	Y		
CRHBP	18-212-200	437	24	Y	T	0.31
CRHBP	18-229-334	438	24	Y	T	0.33
CRHBP	18-230-332	439	24	Y	T	0.32
AVPR1A	18-966-378	440	24	Y	C	0.41
AVPR1A	18-987-308	441	24	Y	A	0.35
AVPR1A	18-1169-118	442	24	Y		
AVPR1A	18-1172-138	443	24	Y	G	0.16
AVPR1A	18-1173-92	444	24	Y	T	0.32
AVPR1A	18-1174-387	445	24	Y	C	0.21
AVPR1A	18-1175-416	446	24	Y	G	0.21
AVPR1A	18-542-146	447	24	Y	G	0.21
5HT1A	8-42-211	448	24	Y	G	0.46
5HT1A	8-45-389	449	24	Y	G	0.01
5HT1A	18-994-270	450	24	Y	C	0.25
5HT1A	18-912-165	451	24	Y	T	0.46
5HT1A	18-991-124	452	24	Y	T	0.45
5HT1A	18-920-219	453	24	Y		
5HT1A	18-911-312	454	24	Y		
5HT1A	99-65963-368	455	24	Y		
5HT1A	99-65966-225	456	24	Y		
5HT1A	99-65968-75	457	24	Y		
5HT1A	99-5069-331	458	24	Y		
5HT1A	99-5070-176	459	24	Y	T	0.02
GABRG2	18-511-348	460	24	Y		
GABRG2	18-523-352	461	24	Y	C	0.49
GABRG2	18-545-478	462	24	Y	G	0.45
GABRG2	18-522-194	463	24	Y	G	0.32
GABRG2	18-524-284	464	24	Y	C	0.36
ADRB1R	18-626-52	465	24	Y	G	0.36
ADRB1R	18-629-189	466	24	Y	T	0.40
ADRB1R	18-1131-71	467	24	Y	T	0.41
ADRB1R	18-534-126	468	24	Y		
ADRB1R	18-596-59	469	24	Y		
ADRB1R	18-597-27	470	24	Y	A	0.20

TABLE 7D (c nt)

GABRA5	18-730-203	471	24	Y	G	0.48
GABRA5	18-734-89	472	24	Y	C	0.48
GABRA5	18-895-321	473	24	Y	A	0.27
GABRA5	18-896-69	474	24	Y		
GABRA5	18-903-58	475	24	Y		
GOLF	18-590-216	476	24	Y	G	0.42
GOLF	18-817-436	477	24	Y	T	0.41
GOLF	18-829-85	478	24	Y	G	0.39
GOLF	18-832-387	479	24	Y		
GOLF	18-833-259	480	24	Y		
GOLF	18-839-271	481	24	Y	C	0.40
GOLF	18-770-194	482	24	Y	T	0.37
GOLF	18-771-302	483	24	Y	G	0.30
GOLF	18-827-53	484	24	Y	G	0.34
GOLF	18-768-318	485	24	Y	G	0.31
GOLF	18-769-26	486	24	Y	A	0.17
SLC6A3	18-709-321	487	24	Y	C	0.46
SLC6A3	18-714-280	488	24	Y	T	0.17
SLC6A3	18-843-271	489	24	Y	C	0.37
SLC6A3	18-850-265	490	24	Y	T	0.34
SLC6A3	18-853-296	491	24	Y	T	0.23
SLC6A3	18-867-331	492	24	Y	C	0.48
SLC6A3	18-877-73	493	24	Y		
SLC6A3	18-856-85	494	24	Y	C	0.42
SLC6A3	18-861-101	495	24	Y	T	0.33
PDE4b	18-635-323	496	24	Y		
PDE4b	18-636-205	497	24	Y	A	0.36
PDE4b	18-649-427	498	24	Y	C	0.46
PDE4b	18-1134-316	499	24	Y	G	0.38
PDE4b	18-633-316	500	24	Y		
COMT	18-489-425	501	24	Y		
COMT	18-492-212	502	24	Y	C	0.41
COMT	18-488-156	503	24	Y	T	0.45
COMT	18-491-266	504	24	Y	T	0.42
COMT	18-497-141	505	24	Y	T	0.43
COMT	18-503-174	506	24	Y	C	0.31
COMT	18-490-95	507	24	Y	A	0.31
NPY1R	18-699-115	508	24	Y		
NPY1R	18-1099-293	509	24	Y		
NPY1R	18-1105-22	510	24	Y		
SLC1	18-562-418	511	24	Y	C	0.49
SLC1	18-564-204	512	24	Y	C	0.496
SEF2-1B	18-1032-262	513	24	Y	C	0.33
SEF2-1B	18-1035-412	514	24	Y	C	0.50
SEF2-1B	18-1036-293	515	24	Y	C	0.34
SEF2-1B	18-1038-95	516	24	Y	T	0.34
SEF2-1B	18-1040-361	517	24	Y	G	0.44
SEF2-1B	18-748-356	518	24	Y	T	0.47
BDNF	18-937-181	519	24	Y	A	0.26
BDNF	18-942-175	520	24	Y	T	0.29
BDNF	18-1213-221	521	24	Y		
BDNF	18-937-147	522	24	Y		
BDNF	18-946-408	523	24	Y	C	0.20
GAP43	18-787-133	524	24	Y	A	0.39
GAP43	18-1149-239	525	24	Y	A	0.49

TABLE 7D (c nt)

GAP43	18-1159-291	526	24	Y	G	0.23
GAP43	18-1135-273	527	24	Y	T	0.43
GAP43	18-1136-108	528	24	Y		
GAP43	18-1147-68	529	24	Y		
GAP43	18-1157-295	530	24	Y		
GAP43	18-802-460	531	24	Y	A	0.32
CLOCK	18-1064-110	532	24	Y	C	0.36
CLOCK	18-1068-327	533	24	Y	T	0.32
CLOCK	18-1069-365	534	24	Y	A	0.23
CLOCK	18-1073-367	535	24	Y	A	0.35
CLOCK	18-1070-272	536	24	Y	T	0.36
CLOCK	18-1057-35	537	24	Y		
CLOCK	18-1062-415	538	24	Y		
CLOCK	18-1082-165	539	24	Y		
CLOCK	18-1080-361	540	24	Y	C	0.18
HSP70	18-506-297	541	24	Y		
HSP70	18-570-38	542	24	Y		

TABLE 8

SEQ ID NO.	BIALLELIC MARKER ID	1ST ALLELE	2ND ALLELE	POSITION RANGE OF PREFERRED SEQUENCE
12	99-28149-118	C	T	[1-478]
13	99-28160-285	A	G	[1-456]
14	99-28171-458	A	G	[1-48],[141-514]
15	99-28173-395	C	T	[1-550]
16	99-32177-113	C	T	[1-466]
17	99-32181-192	C	T	[1-449]
18	99-32193-258	G	T	[1-458]
20	99-28730-351	A	G	[1-452]
21	99-32306-409	G	C	[1-455]
28	99-27097-83	C	T	[1-273]
29	99-27098-162	C	T	[226-421]
32	99-27561-106	A	G	[1-465]
33	99-27562-366	G	T	[1-470]
35	99-27110-301	G	C	[1-455]
37	99-27573-443	G	T	[1-513]
38	99-28732-133	A	G	[1-411]
40	99-28736-399	C	T	[1-453]
41	99-28738-319	C	T	[1-458]
42	99-28739-364	C	T	[1-509]
62	18-132-368	C	T	[1-480]
64	18-12-191	A	C	[1-450]
65	18-11-137	A	G	[1-157],[348-390]
66	18-93-96	G	T	[1-454]
69	18-251-176	C	T	[104-494]
72	18-393-330	G	C	[1-93],[146-479]
73	18-394-402	A	C	[1-21],[119-518]
75	18-284-139	C	T	[146-450]
76	18-285-305	A	G	[1-520]
77	18-289-239	C	T	[1-486]
78	18-291-91	C	T	[1-453]
80	18-194-130	C	T	[1-460]
81	18-198-252	A	G	[1-316],[349-459]
82	18-242-300	A	G	[224-476]
85	99-2409-298	A	G	[117-127],[160-359],[395-711]
86	99-339-54	G	C	[1-247],[293-1514],[1544-2128],[2159-3001]
98	99-32148-315	G	C	[1-24]
108	99-32077-66	A	G	[37-63]
110	99-32376-426	A	G	[235-470]

TABLE 9A

SEQ. ID NO.	BIALLELIC MARKER ID	ORIGINAL ALLELE	ALTERNATIVE ALLELE
1	99-27199-207	T	C
2	99-27207-117	C	T
3	99-27213-53	A	G
4	99-27218-333	T	G
6	99-28109-275	G	A
7	99-28110-75	C	T
8	99-28125-81	A	C
9	99-28134-215	C	T
10	99-28137-96	A	G
11	99-32204-305	G	A
19	99-28722-90	T	C
22	99-27088-246	G	A
23	99-27090-203	G	A
24	99-27091-220	G	A
25	99-27093-145	T	C
26	99-27094-406	T	C
27	99-27096-410	G	A
30	99-27550-48	A	G
31	99-27558-335	C	T
34	16-31-738	C	G
39	99-28735-56	C	T
43	99-27875-185	T	C
44	99-27880-176	T	C
46	99-28753-353	T	C
47	99-28755-206	A	G
48	99-32333-366	T	C
49	16-38-323	A	C
50	99-28484-179	A	T
51	99-30853-364	G	A
52	99-28485-198	G	T
53	99-30858-354	T	C
54	99-32002-313	G	A
55	18-15-366	C	T
56	18-20-174	G	A
59	18-2-192	G	T
60	99-26921-210	G	A
63	18-133-293	C	A
68	16-42-140	A	G
70	18-269-44	A	G
71	16-218-624	G	C
74	16-217-55	A	G
88	10-214-79	C	T
90	99-28779-168	T	C
91	99-28788-300	G	A

TABLE 9B

94	99-32059-169	T	C
95	99-32061-304	A	G
96	99-32065-303	T	G
97	99-32123-118	G	A
99	16-2-76	A	G
101	16-3-199	C	T
102	16-50-197	C	T
103	16-1-59	C	T
104	16-2-187	A	G
105	99-28761-311	A	G
106	99-28771-86	T	C
109	99-32078-466	C	T
111	99-32361-419	T	G
112	16-21-228	G	A
117	16-25-279	G	C
118	16-23-393	G	T
119	16-106-364	T	C
122	16-84-185	T	C
124	16-91-333	G	A
126	16-133-205	A	G
127	16-135-181	T	A
128	16-145-405	C	T
129	16-177-320	A	G

TABLE 10

SEQ ID NO.	BIALLELIC MARKER ID	1ST ALLELE	2ND ALLELE
5	99-28108-233	A	C
36	99-27563-400	A	G
45	99-28747-371	C	T
61	16-215-80	C	T
67	16-115-343	A	C
79	18-186-391	G	T
83	8-15-126	A	G
87	12-254-180	A	G
92	99-32052-262	C	T
93	99-32121-242	A	G
100	16-28-93	A	C
107	99-28791-291	A	G
113	16-22-156	C	T
114	16-23-404	A	G
115	16-24-175	A	C
116	16-25-286	C	T
120	16-16-285	C	T
121	16-17-121	C	T
123	16-87-74	A	G
125	16-128-142	C	G
130	16-4-354	C	T

TABLE 11

SEQ. ID NO.	POSITION RANGE OF PREFERRED SEQUENCE
1	[103-147]
7	[1-25]
8	[508-518]
9	[398-432]
10	[295-364]
11	[301-342]
23	[246-287]
25	[369-413]
30	[126-153],[182-468]
31	[271-313],[443-452]
34	[408-461]
39	[147-235],[438-457]
43	[498-549]
46	[432-448]
49	[263-320]
54	[472-489]
59	[280-321]
63	[486-505]
71	[258-437],[669-927]
74	[90-165]
79	[1-82],[150-191]
83	[1-16],[144-498],[620-800],[1300-1366], [1823-1908],[2336-2365],[2398-3001]
88	[1-1297],[1998-2689],[2895-2965]
92	[255-348],[493-499]
93	[445-467]
94	[1-16],[396-438]
96	[246-288]
97	[1-91],[420-541]
111	[443-457]
121	[130-181]
122	[160-399]
123	[144-145],[351-435]
128	[283-551]
129	[375-458]

TABLE 12

SEQ ID NO.	POSITION RANGE OF MICROSEQUENCING PRIMERS	COMPLEMENTARY POSITION RANGE OF MICROSEQUENCING PRIMERS
1	188-206*	208-227
2	98-116*	118-136*
3	34-52*	54-73
4	314-332*	334-353
5	214-232*	234-253
6	255-274	276-295
7	54-73	75-94
8	62-80*	82-101
9	196-214*	216-235
10	77-95*	97-116
11	285-304	306-325
12	98-117	119-137*
13	266-284*	286-305
14	438-456*	458-477
15	375-394	396-414*
16	93-112	114-132*
17	172-191	193-212
18	238-256*	258-277
19	71-89*	91-110
20	332-350*	352-370*
21	387-406	408-427
22	226-245	247-266
23	185-203*	205-224
24	201-220	222-241
25	125-144	146-165
26	387-405*	407-426
27	390-409	411-430
28	63-82	84-103
29	142-161	163-182
30	28-47	49-67*
31	316-334*	336-355
32	87-105*	107-126
33	344-363	365-384
34	715-737*	739-761*
35	281-299*	301-319*
36	381-399*	401-420
37	423-442	444-462*
38	114-132*	134-153
39	36-55	57-76
40	379-398	400-418*
41	299-318	320-338*
42	345-363*	365-384
43	165-184	186-204*

TABLE 12 (cont)

44	156-175	177-195*
45	353-372	374-392*
46	332-351	353-371*
47	187-206	208-226*
48	346-365	367-386
49	300-322*	324-346*
50	160-178*	180-199
51	345-363*	365-384
52	179-197*	199-217*
53	334-353	355-373*
54	292-310*	312-331
55	347-365*	367-385*
56	155-173*	175-194
57	158-177	179-197*
58	375-394	396-414*
59	172-191	193-211*
60	192-210*	212-231
61	231-249*	251-270
62	348-367	369-387*
63	273-291*	293-311*
64	172-190*	192-210*
65	118-137	139-157*
66	77-95*	97-115*
67	320-342*	344-366*
68	121-139*	141-163*
69	157-175*	177-196
70	25-43*	45-64
71	601-623*	625-644
72	311-329*	331-349*
73	382-401	403-421*
74	32-54*	56-74*
75	120-138*	140-159
76	286-304*	306-325
77	219-238	240-258*
78	71-90	92-110*
79	371-390	392-410*
80	110-129	131-149*
81	233-251*	253-272
82	280-298*	300-319
83	1481-1500	1502-1520*
84	1481-1500	1502-1520*
85	408-427	429-447*
86	1482-1500*	1502-1521
87	292-310*	312-331
88	1482-1500*	1502-1521
89	1482-1500*	1502-1521
90	149-167*	169-187*

TABLE 12 (cont.)

91	281-299*	301-320
92	244-262*	264-282*
93	225-243*	245-264
94	149-168	170-189
95	285-303*	305-324
96	284-302*	304-322*
97	99-117*	119-138
98	294-313	315-333*
99	76-94*	96-114*
100	101-119*	121-143*
101	323-341*	343-361*
102	178-196*	198-216*
103	158-180*	182-200*
104	183-205*	207-225*
105	292-310*	312-331
106	67-85*	87-106
107	272-290*	292-311
108	47-65*	67-85*
109	448-466*	468-486*
110	407-425*	427-446
111	400-419	421-439*
112	205-227*	229-251*
113	133-155*	157-175*
114	381-403*	405-427*
115	156-174*	176-194*
116	267-285*	287-305*
117	260-278*	280-298*
118	370-392*	394-416*
119	345-363*	365-383*
120	265-284	286-304*
121	102-120*	122-140*
122	162-184*	186-208*
123	51-73*	75-97*
124	310-332*	334-356*
125	123-141*	143-161*
126	222-244*	246-268*
127	209-231*	233-251*
128	436-454*	456-474*
129	297-319*	321-343*
130	335-353*	355-373*
261	343-361*	363-382
262	68-87	89-107*
263	316-334*	336-355
264	48-66*	68-87

TABLE 12 (c nt.)

265	295-313*	315-333*
266	302-321	323-341*
267	296-314*	316-334*
268	327-345*	347-366
269	244-262*	264-282*
270	1482-1500*	1502-1521
271	115-134	136-154*
272	400-418*	420-439
273	271-289*	291-309*
274	237-255*	257-276
275	48-67	69-87*
276	246-265	267-285*
277	287-305*	307-326
278	1482-1500*	1502-1521
279	1482-1500*	1502-1521
280	1482-1500*	1502-1521
281	225-244	246-264*
282	271-290	292-310*
283	327-345*	347-366
284	387-405*	407-426
285	319-337*	339-357*
286	91-109*	111-130
287	85-103*	105-124
288	12-30*	32-50*
289	323-341*	343-362
290	121-139*	141-160
291	160-178*	180-199
292	304-322*	324-343
293	50-69	71-89*
294	422-441	443-461*
295	46-64*	66-85
296	181-199*	201-220
297	314-333	335-353*
298	313-331*	333-352
299	358-377	379-397*
300	288-306*	308-327
301	99-117*	119-138
302	119-137*	139-158
303	72-91	93-111*
304	368-386*	388-407
305	397-415*	417-436
306	127-145*	147-165*
307	1481-1500	1482-1500*
308	1481-1500	1502-1520*
309	250-269	271-289*
310	146-164*	166-184*
311	105-123*	125-144
312	199-218	220-238*
313	293-311*	313-331*
314	349-367*	369-388
315	206-224*	226-245
316	56-74*	76-95
317	1482-1500*	1502-1521

TABLE. 12 (c nt.)

318	156-174*	176-194*
319	328-347	349-367*
320	332-351	353-371*
321	461-479*	481-500
322	175-193*	195-214
323	265-283*	285-304
324	33-51*	53-72
325	169-188	190-208*
326	51-70	72-90*
327	106-125	127-145*
328	40-58*	60-79
329	7-26	28-46*
330	184-202*	204-223
331	70-88*	90-108*
332	302-320*	322-341
333	50-68*	70-89
334	39-57*	59-77*
335	197-215*	217-236
336	414-432*	434-452*
337	66-84*	86-105
338	368-386*	388-407
339	239-258	260-278*
340	252-270*	272-291
341	175-193*	195-213*
342	283-301*	303-321*
343	33-52	54-72*
344	299-317*	319-338
345	6-25	27-45*
346	300-319	321-339*
347	262-280*	282-300*
348	252-270*	272-290*
349	246-264*	266-284*
350	277-295*	297-315*
351	313-331*	333-352
352	54-72*	74-93
353	66-84*	86-104*
354	82-100*	102-120*
355	304-322*	324-343
356	186-204*	206-225
357	408-426*	428-447
358	297-315*	317-336
359	297-315*	317-336
360	406-424*	426-445
361	193-211*	213-231*
362	137-155*	157-175*
363	247-265*	267-285*
364	122-140*	142-160*
365	155-173*	175-194
366	75-94	96-114*
367	95-113*	115-134
368	274-292*	294-312*
369	3-21*	23-42
370	399-417*	419-438

TABLE 12 (cont.)

371	185-203*	205-224
372	242-260*	262-281
373	393-411*	413-431*
374	274-292*	294-313
375	75-94	96-114*
376	342-360*	362-381
377	336-355	357-375*
378	160-178*	180-198*
379	156-174*	176-195
380	202-220*	222-241
381	126-144*	146-165
382	387-406	408-426*
383	113-132	134-152*
384	220-238*	240-259
385	272-290*	292-311
386	254-272*	274-293
387	89-107*	109-128
388	49-67*	69-88
389	276-294*	296-315
390	440-458*	460-479
391	89-108	110-128*
392	307-326	328-346*
393	345-364	366-384*
394	348-366*	368-387
395	253-271*	273-292
396	16-34*	36-54*
397	396-414*	416-435
398	146-164*	166-185
399	343-362	364-382*
400	278-296*	298-317
401	19-37*	39-58

TABLE 13

SEQ ID NO.	POSITION RANGE OF AMPLIFICATION PRIMERS	COMPLEMENTARY POSITION RANGE OF AMPLIFICATION PRIMERS
1	1-20	431-450
2	1-21	432-452
3	1-18	446-464
4	1-20	528-546
5	1-18	394-413
6	1-18	434-454
7	1-18	530-549
8	1-18	500-518
9	1-20	453-472
10	1-18	529-546
11	1-20	384-401
12	1-19	459-478
13	1-19	439-456
14	1-18	494-514
15	1-18	532-550
16	1-19	446-466
17	1-18	432-449
18	1-19	438-458
19	1-17	429-449
20	1-18	435-451
21	1-18	437-454
22	1-19	510-527
23	1-20	431-451
24	1-20	455-473
25	1-20	453-472
26	1-20	436-455
27	1-18	432-450
28	1-18	486-504
29	1-19	404-421
30	1-19	448-468
31	1-18	432-452
32	1-19	446-465
33	1-19	450-470
34	1-25	975-1003
35	1-18	438-455
36	1-18	526-546
37	1-18	496-513
38	1-19	391-410
39	1-19	438-456
40	1-20	434-452
41	1-18	441-457
42	1-17	489-508
43	1-18	531-549

TABLE 13 (c nt)

44	1-18	444-462
45	1-19	478-496
46	1-20	427-447
47	1-17	452-470
48	1-20	520-540
49	1-28	389-416
50	1-17	488-505
51	1-17	465-485
52	1-17	449-466
53	1-17	456-473
54	1-19	472-488
55	1-20	507-525
56	1-17	408-425
57	1-19	437-457
58	1-19	456-473
59	1-19	450-468
60	1-21	431-451
61	171-188	284-303
62	1-18	461-479
63	1-17	487-504
64	1-18	431-449
65	1-18	516-535
66	1-17	436-453
67	1-24	533-553
68	1-18	154-171
69	1-19	474-493
70	1-17	457-477
71	1-22	906-927
72	1-18	459-479
73	1-20	500-518
74	1-20	568-587
75	1-17	430-449
76	1-17	499-519
77	1-17	466-485
78	1-17	432-452
79	1-18	442-459
80	1-17	439-459
81	1-17	438-458
82	1-17	455-475
83	1376-1395	1792-1810
84	1130-1148	1534-1552
85	131-148	560-580
86	1448-1467	1883-1902
87	132-152	586-603
88	1225-1244	1747-1764
89	1414-1430	1759-1775
90	1-17	390-409

TABLE 13 (cont)

91	1-17	458-478
92	1-18	478-498
93	1-17	448-466
94	1-17	448-468
95	1-19	430-449
96	1-17	469-486
97	1-20	520-540
98	1-18	428-448
99	20-39	240-260
100	28-47	354-374
101	143-162	374-393
102	1-20	227-245
103	123-142	290-309
104	20-39	240-260
105	1-18	446-465
106	1-18	444-461
107	1-18	432-451
108	1-18	471-488
109	1-17	470-488
110	1-18	449-469
111	1-18	442-456
112	1-25	399-424
113	1-24	458-481
114	1-26	455-478
115	1-22	405-428
116	1-22	412-433
117	1-22	412-433
118	1-26	455-478
119	1-22	723-742
120	1-19	516-535
121	1-19	508-529
122	1-22	525-540
123	1-22	504-525
124	1-19	641-665
125	1-20	308-327
126	41-59	472-490
127	52-71	482-501
128	51-69	523-540
129	1-22	472-492
130	1-20	740-759
261	1-21	482-502
262	1-21	438-457
263	1-20	482-502
264	1-19	441-461
265	1-18	428-448
266	1-19	409-426
267	1-19	403-422
268	1-19	401-419
269	1-18	478-498

TABLE 13(cont.)

270	1211-1229	1588-1606
271	1-18	448-465
272	1-20	507-527
273	1-18	434-451
274	1-17	466-486
275	1-19	436-453
276	1-18	452-471
277	1-18	450-468
278	1156-1173	1652-1672
279	1149-1166	1591-1608
280	1170-1187	1635-1652
281	1-18	440-460
282	1-18	433-453
283	1-17	538-558
284	1-17	450-467
285	1-18	439-456
286	1-18	439-456
287	1-20	431-451
288	1-20	431-451
289	1-20	431-451
290	1-19	458-476
291	1-18	497-517
292	1-18	419-436
293	1-17	487-504
294	1-17	443-463
295	1-18	438-455
296	1-18	463-483
297	1-20	464-484
298	1-19	439-456
299	1-18	458-478
300	1-18	443-463
301	1-18	442-460
302	1-18	457-475
303	1-18	515-533
304	1-18	443-463
305	1-19	434-451
306	1-21	430-450
307	1263-1281	1694-1711
308	1114-1133	1516-1533
309	1-19	481-498
310	1-18	447-467
311	1-20	444-463
312	1-19	534-551
313	1-19	437-457
314	1-19	459-477
315	1-19	486-502
316	1-18	432-452
317	1171-1189	1702-1719
318	1-18	476-493
319	1-20	430-450
320	1-20	455-475
321	1-21	489-509
322	1-21	457-477

TABLE 13 (cont.)

323	1-19	430-450
324	1-21	485-505
325	1-20	466-486
326	1-21	437-457
327	1-19	497-517
328	1-20	495-514
329	1-20	450-470
330	1-18	456-476
331	1-20	433-453
332	1-19	435-455
333	1-19	544-561
334	1-18	442-459
335	1-20	434-453
336	1-21	514-534
337	1-20	464-483
338	1-21	580-597
339	1-18	441-461
340	1-20	430-450
341	1-18	479-496
342	1-21	461-481
343	1-21	429-449
344	1-21	409-429
345	1-21	430-450
346	1-20	433-453
347	1-19	495-514
348	1-20	433-451
349	1-20	443-460
350	1-20	440-459
351	1-18	445-463
352	1-18	432-449
353	1-19	430-450
354	1-19	462-480
355	1-21	445-465
356	1-18	473-493
357	1-18	570-589
358	1-20	412-432
359	1-20	412-432
360	1-20	515-533
361	1-20	465-485
362	1-18	550-570
363	1-20	430-450
364	1-21	416-435
365	1-20	455-475
366	1-20	481-501
367	1-21	428-448
368	1-19	459-478
369	1-18	456-476
370	1-20	475-495
371	1-18	456-476
372	1-20	447-467
373	1-21	447-466
374	1-18	467-484
375	1-18	466-484

TABLE 13 (c nt.)

376	1-19	437-455
377	1-20	383-403
378	1-20	428-448
379	1-20	431-449
380	1-18	434-452
381	1-20	428-448
382	1-19	453-473
383	1-21	480-497
384	1-21	532-552
385	1-20	480-500
386	1-18	429-449
387	1-21	433-450
388	1-21	430-450
389	1-21	576-595
390	1-21	549-569
391	1-18	438-455
392	1-19	496-516
393	1-19	455-472
394	1-20	441-458
395	1-19	455-475
396	1-18	449-469
397	1-18	557-577
398	1-18	433-453
399	1-18	475-494
400	1-20	491-511
401	1-21	380-400

TABLE 14

SEQ ID NO.	POSITION RANGE OF PROBES
1	195-219
2	105-129
3	41-65
4	321-345
5	221-245
6	263-287
7	62-86
8	69-93
9	203-227
10	84-108
11	293-317
12	106-130
13	273-297
14	445-469
15	383-407
16	101-125
17	180-204
18	245-269
19	78-102
20	339-363
21	395-419
22	234-258
23	192-216
24	209-233
25	133-157
26	394-418
27	398-422
28	71-95
29	150-174
30	36-60
31	323-347
32	94-118
33	352-376
34	726-750
35	288-312
36	388-412
37	431-455
38	121-145
39	44-68
40	387-411
41	307-331
42	352-376
43	173-197
44	164-188
45	361-385

TABLE 14 (ont.)

TABLE 14 (cont.)

46	340-364
47	195-219
48	354-378
49	311-335
50	167-191
51	352-376
52	186-210
53	342-366
54	299-323
55	354-378
56	162-186
57	166-190
58	383-407
59	180-204
60	199-223
61	238-262
62	356-380
63	280-304
64	179-203
65	126-150
66	84-108
67	331-355
68	128-152
69	164-188
70	32-56
71	612-636
72	318-342
73	390-414
74	43-67
75	127-151
76	293-317
77	227-251
78	79-103
79	379-403
80	118-142
81	240-264
82	287-311
83	1489-1513
84	1489-1513
85	416-440
86	1489-1513
87	299-323
88	1489-1513
89	1489-1513
90	156-180
91	288-312
92	251-275
93	232-256
94	157-181
95	292-316
96	291-315
97	106-130

TABLE 14 (cont.)

98	302-326
99	83-107
100	108-132
101	330-354
102	185-209
103	169-193
104	194-218
105	299-323
106	74-98
107	279-303
108	54-78
109	455-479
110	414-438
111	408-432
112	216-240
113	144-168
114	392-416
115	163-187
116	274-298
117	267-291
118	381-405
119	352-376
120	273-297
121	109-133
122	173-197
123	62-86
124	321-345
125	130-154
126	233-257
127	220-244
128	443-467
129	308-332
130	342-366
261	350-374
262	76-100
263	323-347
264	55-79
265	302-326
266	310-334
267	303-327
268	334-358
269	251-275
270	1489-1513
271	123-147
272	407-431
273	278-302
274	244-268
275	56-80
276	254-278
277	294-318
278	1489-1513
279	1489-1513
280	1489-1513

TABLE 14 (cont.)

281	233-257
282	279-303
283	334-358
284	394-418
285	326-350
286	98-122
287	92-116
288	19-43
289	330-354
290	128-152
291	167-191
292	311-335
293	58-82
294	430-454
295	53-77
296	188-212
297	322-346
298	320-344
299	366-390
300	295-319
301	106-130
302	126-150
303	80-104
304	375-399
305	404-428
306	134-158
307	1489-1513
308	1489-1513
309	258-282
310	153-177
311	112-136
312	207-231
313	300-324
314	356-380
315	213-237
316	63-87
317	1489-1513
318	163-187
319	336-360
320	340-364
321	468-492
322	182-206
323	272-296
324	40-64
325	177-201
326	59-83
327	114-138
328	47-71
329	15-39
330	191-215
331	77-101
332	309-333
333	57-81

TABLE 14 (cont.)

334	46-70
335	204-228
336	421-445
337	73-97
338	375-399
339	247-271
340	259-283
341	182-206
342	290-314
343	41-65
344	306-330
345	14-38
346	308-332
347	269-293
348	259-283
349	253-277
350	284-308
351	320-344
352	61-85
353	73-97
354	89-113
355	311-335
356	193-217
357	415-439
358	304-328
359	304-328
360	413-437
361	200-224
362	144-168
363	254-278
364	129-153
365	162-186
366	83-107
367	102-126
368	281-305
369	10-34
370	406-430
371	192-216
372	249-273
373	400-424
374	281-305
375	83-107
376	349-373
377	344-368
378	167-191
379	163-187
380	209-233
381	133-157
382	395-419
383	121-145
384	227-251
385	279-303
386	261-285

387	96-120
388	56-80
389	283-307
390	447-471
391	97-121
392	315-339
393	353-377
394	355-379
395	260-284
396	23-47
397	403-427
398	153-177
399	351-375
400	285-309
401	26-50

TABLE 15

Gene	Marker	Allele 1	Allele 2	Allele 1 Frequencies			Genotype Frequencies							Numbers	
				Cases	Controls	p-value	Cases			Controls			p-value	Cases	Controls
							11	12	22	11	12	22			
5HTR6	99-27199/207	C	T	0.72	0.70	0.71	0.51	0.41	0.08	0.48	0.45	0.07	0.83	140	94
5HTR6	99-27207/117	C	T	0.48	0.37	0.02	0.24	0.47	0.29	0.13	0.47	0.40	0.06	139	93
5HTR6	99-28110/75	C	T	0.64	0.52	0.01	0.41	0.45	0.14	0.27	0.51	0.22	0.05	139	90
5HTR6	99-28134/215	C	T	0.51	0.63	0.01	0.26	0.50	0.24	0.42	0.43	0.15	0.04	136	93
5HTR7	99-28160/285	A	G	0.49	0.51	0.73	0.27	0.44	0.29	0.24	0.53	0.23	0.37	140	91
5HTR7	99-28171/458	A	G	0.34	0.38	0.44	0.13	0.42	0.45	0.17	0.42	0.41	0.74	136	90
5HTR7	99-28173/395	C	T	0.51	0.55	0.33	0.27	0.47	0.26	0.27	0.57	0.16	0.18	137	94
5HTR7	99-32181/192	C	T	0.28	0.38	0.02	0.09	0.38	0.54	0.13	0.51	0.36	0.04	136	85
5HTR7	99-32193/258	G	T	0.70	0.65	0.26	0.53	0.34	0.12	0.41	0.48	0.11	0.12	137	85
CHRNA7	99-28722/90	C	T	0.27	0.32	0.27	0.00	0.54	0.46	0.01	0.62	0.37	0.22	140	94
CHRNA7	99-28730/351	A	G	0.32	0.38	0.17	0.04	0.55	0.41	0.03	0.69	0.28	0.09	139	94
CHRNA7	99-32306/409	C	G	0.62	0.67	0.31	0.26	0.72	0.01	0.34	0.66	0.00	0.25	137	91
CRFR1	99-27088/246	A	G	0.42	0.38	0.31	0.20	0.44	0.36	0.15	0.45	0.40	0.57	138	93
CRFR1	99-27091/220	A	G	0.44	0.49	0.32	0.19	0.50	0.31	0.24	0.50	0.26	0.59	137	92
CRFR1	99-27097/83	C	T	0.55	0.53	0.73	0.34	0.41	0.24	0.26	0.56	0.19	0.11	140	90
CRFR1	99-27550/48	A	G	0.20	0.26	0.17	0.04	0.32	0.64	0.07	0.38	0.55	0.38	138	91
MLR	19-26/204/A23	C	G	0.55	0.53	0.70	0.28	0.55	0.18	0.28	0.50	0.22	0.69	137	92
MLR	99-27563/400	A	G	0.46	0.44	0.79	0.19	0.54	0.28	0.21	0.46	0.33	0.54	138	89
MLR	99-28732/133	A	G	0.39	0.31	0.09	0.16	0.47	0.37	0.09	0.44	0.47	0.23	137	86
MLR	99-28735/56	C	T	0.71	0.70	0.87	0.50	0.41	0.09	0.50	0.40	0.10	0.93	136	88
CRFR2	99-27875/185	C	T	0.36	0.40	0.30	0.16	0.39	0.45	0.16	0.49	0.35	0.25	140	94
CRFR2	99-27880/176	C	T	0.53	0.56	0.47	0.30	0.46	0.24	0.32	0.47	0.20	0.76	139	93
CRFR2	99-28747/371	C	T	0.39	0.44	0.34	0.16	0.47	0.37	0.23	0.42	0.35	0.39	138	91
CRFR2	99-28755/206	A	G	0.56	0.59	0.65	0.32	0.50	0.19	0.37	0.44	0.19	0.68	139	93

TABLE 15 (cont.)

GRL	99-30853/364	A	G	0.56	0.58	0.62	0.35	0.41	0.24	0.39	0.39	0.23	0.87	139	93
GRL	99-30858/354	C	T	0.76	0.84	0.04	0.59	0.34	0.06	0.71	0.27	0.02	0.13	140	89
GRL	99-28485/198	G	T	0.19	0.20	0.68	0.04	0.30	0.66	0.01	0.38	0.61	0.25	134	89
GRL	99-32002/313	A	G	0.48	0.48	0.97	0.23	0.51	0.26	0.16	0.65	0.19	0.10	140	94
GRL	18-20/174	A	G	0.76	0.73	0.38	0.57	0.39	0.04	0.53	0.39	0.07	0.55	140	94
GRL	18-31/178	C	T	0.30	0.35	0.21	0.11	0.38	0.51	0.15	0.40	0.45	0.49	140	94
GRL	18-38/395	A	T	0.59	0.63	0.38	0.36	0.46	0.18	0.36	0.54	0.10	0.17	140	94
MAO A/B	18-2/192	G	T	0.70	0.68	0.66	0.56	0.26	0.17	0.56	0.24	0.20	0.78	140	93
MAO A/B	19-25/407	C	T	0.67	0.67	1.00	0.54	0.26	0.20	0.54	0.24	0.21	0.96	138	90
MAO A/B	99-26921/210	A	G	0.48	0.52	0.47	0.33	0.30	0.37	0.31	0.41	0.28	0.20	139	93
MAO A/B	18-133/293	A	C	0.30	0.27	0.46	0.16	0.29	0.55	0.14	0.26	0.60	0.77	140	92
NET	19-56/140	A	G	0.34	0.27	0.11	0.12	0.44	0.44	0.02	0.49	0.48	0.03	134	93
NET	19-14/241	C	T	0.35	0.32	0.64	0.12	0.46	0.43	0.10	0.45	0.45	0.88	136	91
NET	19-28/136	A	C	0.68	0.56	0.01	0.44	0.48	0.08	0.28	0.57	0.15	0.02	140	93
NET	19-44/251	C	T	0.24	0.21	0.40	0.09	0.31	0.60	0.02	0.38	0.60	0.11	140	93
NET	99-28788/300	A	G	0.49	0.47	0.65	0.22	0.54	0.24	0.21	0.51	0.28	0.85	140	94
NET	99-32061/304	A	G	0.41	0.39	0.71	0.16	0.51	0.34	0.14	0.51	0.35	0.92	140	94
NET	99-32121/242	A	G	0.61	0.52	0.07	0.36	0.49	0.15	0.29	0.47	0.24	0.16	140	94
TACR1	99-28761/311	A	G	0.25	0.22	0.38	0.07	0.36	0.56	0.01	0.41	0.57	0.09	140	94
TACR1	99-28771/86	C	T	0.48	0.52	0.33	0.24	0.46	0.29	0.23	0.57	0.19	0.16	140	94
TACR1	99-28791/291	A	G	0.28	0.26	0.58	0.09	0.39	0.53	0.04	0.43	0.53	0.42	140	94
TACR1	99-32361/419	G	T	0.53	0.52	0.79	0.29	0.49	0.23	0.26	0.52	0.22	0.84	140	94
DRD3	8-15/126	A	G	0.64	0.70	0.16	0.43	0.42	0.15	0.47	0.47	0.06	0.13	140	94
DRD3	8-19/372	A	G	0.23	0.28	0.22	0.09	0.29	0.62	0.03	0.50	0.47	0.004	140	94
DRD3	99-2409/298	A	G	0.36	0.40	0.30	0.14	0.44	0.42	0.11	0.60	0.30	0.07	140	94
DRD3	99-339/54	C	G	0.55	0.54	0.93	0.34	0.42	0.24	0.28	0.53	0.19	0.25	140	94
CYP3A4	12-254/180	A	G	0.43	0.54	0.02	0.19	0.47	0.34	0.32	0.45	0.23	0.06	140	94
CYP3A4	10-214/279	C	T	0.14	0.12	0.55	0.02	0.23	0.75	0.02	0.19	0.79	0.79	140	94
CYP3A4	10-217/91	C	T	0.89	0.93	0.16	0.79	0.19	0.02	0.86	0.13	0.01	0.39	140	94

TABLE 15 (cont)

5HTT	18-194/130	C	T	0.54	0.52	0.67	0.31	0.44	0.24	0.20	0.63	0.17	0.02	140	94
5HTT	18-198/252	A	G	0.52	0.49	0.57	0.29	0.47	0.24	0.20	0.59	0.21	0.20	140	94
5HTT	18-242/300	A	G	0.46	0.54	0.09	0.24	0.44	0.32	0.27	0.54	0.19	0.09	140	94
5HTT	18-186/391	G	T	0.50	0.53	0.52	0.27	0.45	0.28	0.26	0.54	0.20	0.31	140	94
DRD2	19-23/215	A	G	0.21	0.16	0.24	0.05	0.31	0.64	0.04	0.24	0.72	0.44	139	93
DRD2	19-5/377	A	G	0.55	0.53	0.76	0.30	0.50	0.20	0.26	0.55	0.19	0.78	139	93
DRD2	19-6/171	A	C	0.22	0.16	0.15	0.06	0.32	0.62	0.04	0.24	0.72	0.34	138	92
DRD2	19-7/275	C	T	0.70	0.63	0.12	0.49	0.43	0.08	0.38	0.52	0.11	0.24	135	93
DRD2	19-4/118	C	T	0.46	0.45	0.79	0.21	0.50	0.29	0.17	0.56	0.27	0.63	140	93
Gbeta3	19-58/162	C	T	0.70	0.62	0.07	0.49	0.42	0.09	0.36	0.51	0.13	0.16	139	91
Gbeta3	19-9/45	C	T	0.71	0.64	0.12	0.49	0.43	0.08	0.40	0.47	0.13	0.28	139	92
Gbeta3	19-88/185	C	T	0.43	0.40	0.50	0.24	0.38	0.38	0.25	0.29	0.46	0.41	133	92
Gbeta3	19-22/74	A	G	0.33	0.34	0.79	0.09	0.49	0.43	0.14	0.40	0.46	0.30	140	94
Gbeta3	20-205/302	C	T	0.94	0.99	0.002	0.88	0.12	0.00	0.99	0.01	0.00	0.002	140	94
WFS1	19-19/174	C	T	0.43	0.36	0.15	0.17	0.51	0.32	0.11	0.51	0.39	0.30	139	94
WFS1	19-16/127	C	G	0.33	0.27	0.19	0.09	0.49	0.43	0.05	0.44	0.51	0.38	140	94
WFS1	19-17/188	A	G	0.58	0.66	0.07	0.32	0.51	0.16	0.43	0.46	0.11	0.18	140	93
WFS1	19-18/310	A	G	0.14	0.07	0.04	0.02	0.23	0.75	0.01	0.13	0.86	0.12	140	94
TH	19-15/324	A	C	0.69	0.76	0.10	0.51	0.37	0.12	0.60	0.32	0.08	0.31	132	91
TH	18-251/176	C	T	0.58	0.57	0.87	0.36	0.44	0.20	0.29	0.57	0.14	0.11	140	94
TH	18-269/44	A	G	0.36	0.38	0.71	0.11	0.51	0.39	0.15	0.46	0.39	0.58	140	94
5HTR2c	18-12/191	A	C	0.15	0.14	0.73	0.66	0.17	0.77	0.03	0.21	0.76	0.40	137	94

TABLE 16A

OMNIBUS LR RANK of HAPLOTYPES

OMNIBUS LR Rank of 4-locus combinations

Gene	Cases	Controls	Marker1	Marker2	Marker3	Marker4	p-value	-log p-value
NET	Genset	Argent	19-56/140	99-28788/300	99-32061/304	99-32121/242	0.001	3.00
NET	Genset	Argent	19-28/136	99-28788/300	99-32061/304	99-32121/242	0.001	3.00
Gbeta3	Genset	Argent	19-58/162	19-9/45	19-22/74	20-205/302	0.002	2.70
NET	Genset	Argent	16-3/199	19-28/136	99-32061/304	99-32121/242	0.002	2.70
Gbeta3	Genset	Argent	19-58/162	19-9/45	19-88/185	20-205/302	0.003	2.52
NET	Genset	Argent	19-56/140	19-28/136	99-32061/304	99-32121/242	0.003	2.52
NET	Genset	Argent	19-56/140	16-3/199	99-32061/304	99-32121/242	0.005	2.30
NET	Genset	Argent	19-28/136	16-50/196	99-32061/304	99-32121/242	0.010	2.00
5HTT	Genset	Argent	18-186/391	18-194/130	18-198/252	18-242/300	0.013	1.89
NET	Genset	Argent	19-56/140	16-50/196	99-32061/304	99-32121/242	0.014	1.85

92 total combinations

TABLE 16B

OMNIBUS LR Rank of 3-locus combinations

Gene	Cases	Controls	Marker1	Marker2	Marker3	Marker4	p-value	-log p-value
Gbeta3	Genset	Argent	19-58/162	19-9/45	20-205/302	0	0.0001	4.00
WFS1	Genset	Argent	19-19/174	19-17/188	19-18/310	0	0.0003	3.55
NET	Genset	Argent	16-3/199	19-28/136	16-50/196	0	0.0009	3.05
NET	Genset	Argent	19-28/136	99-32061/304	99-32121/242	0	0.001	3.00
WFS1	Genset	Argent	19-19/174	19-16/127	19-17/188	0	0.001	3.00
Gbeta3	Genset	Argent	19-58/162	19-22/74	20-205/302	0	0.002	2.70
NET	Genset	Argent	99-28788/300	99-32061/304	99-32121/242	0	0.002	2.70
NET	Genset	Argent	19-56/140	99-32061/304	99-32121/242	0	0.004	2.40
NET	Genset	Argent	16-3/199	19-28/136	16-50/196	0	0.004	2.40
Gbeta3	Genset	Argent	19-9/45	19-22/74	20-205/302	0	0.006	2.22

136 total combinations

TABLE 16 C

OMNIBUS LR Rank of 2-locus combinations

Gene	Cases	Controls	Marker1	Marker2	Marker3	Marker4	p-value	-log p-value
Gbeta3	Genset	Argent	19-58/162	20-205/302	0	0	0.00003	4.52
WFS1	Genset	Argent	19-19/174	19-17/188	0	0	0.00021	3.68
Gbeta3	Genset	Argent	19-9/45	20-205/302	0	0	0.001	3.00
NET	Genset	Argent	99-32061/304	99-32121/242	0	0	0.003	2.52
Gbeta3	Genset	Argent	19-22/74	20-205/302	0	0	0.006	2.22
5HTR7	Genset	Argent	99-28106/185	99-32181/192	0	0	0.007	2.15
DRD3	Genset	Argent	8-15/126	99-2409/298	0	0	0.014	1.87
Gbeta3	Genset	Argent	19-88/185	20-205/302	0	0	0.014	1.85
5HTR7	Genset	Argent	99-32181/192	99-32193/258	0	0	0.015	1.82
5HTT	Genset	Argent	18-186/394	18-242/300	0	0	0.016	1.80

125 total combinations

TABLE 17A

Rank of Permutation Tests for Individual Haplotypes

4-Locus Combination

					EM Methods						Permutation Test			Omnibus LR		
Gene	Marker1	Marker2	Marker3	Marker4	Haplotype	Case	Control	Difference	Odds Ratio	Chi-Square	number	p-value	-log p-value	LR statistic	p-value	-log p-value
NET	19-56/140	19-14/241	19-28/136	16-50/196	GTCT	0.28	0.41	-0.13	0.56	4.02	100000	0.002	2.61	19.59	0.075	1.12
GRL	18-20/174	99-32002/313	18-31/178	18-38/395	AGCA	0.06	0.17	-0.12	0.28	8.19	100000	0.003	2.57	16.09	0.360	0.44
GRL	18-20/174	18-31/178	18-38/395	99-30858/354	GTAC	0.11	0.22	-0.10	0.46	4.56	100000	0.003	2.55	25.70	0.045	1.35
GRL	18-20/174	18-38/395	99-30853/364	99-30858/354	GAAC	0.03	0.12	-0.09	0.26	6.37	1000	0.003	2.52	21.83	0.078	1.11
GRL	99-32002/313	18-31/178	18-38/395	99-30858/354	GCAC	0.06	0.14	-0.09	0.36	4.91	1000	0.003	2.52	20.00	0.159	0.80
GRL	99-32002/313	18-31/178	18-38/395	99-30858/354	GTAT	0.13	0.05	0.09	3.26	4.77	1000	0.003	2.52	20.00	0.159	0.80
5HTT	18-186/391	18-194/130	18-198/252	18-242/300	GCAA	0.02	0.08	-0.06	0.27	4.13	1000	0.003	2.52	25.66	0.013	1.89
NET	19-28/136	99-28788/300	99-32061/304	99-32121/242	AGGG	0.01	0.05	-0.04	0.14	4.28	1000	0.004	2.40	31.97	0.001	3.00
GRL	99-32002/313	18-31/178	99-30853/364	99-30858/354	GTAT	0.18	0.08	0.10	2.42	4.20	1000	0.004	2.38	16.19	0.314	0.50
5HTR7	99-28106/185	99-28171/458	99-28173/395	99-32181/192	AACC	0.27	0.38	-0.11	0.60	2.90	1000	0.005	2.30	13.67	0.034	1.47
NET	19-14/241	19-28/136	16-50/196	99-32061/304	TCTG	0.17	0.28	-0.11	0.53	3.88	1000	0.006	2.22	20.75	0.084	1.08
GRL	18-31/178	18-38/395	99-30853/364	99-30858/354	TAAT	0.14	0.05	0.09	3.06	4.68	1000	0.006	2.22	22.38	0.090	1.05
DRD3	8-15/126	8-19/372	99-2409/298	99-339/54	GGAG	0.06	0.01	0.06	9.29	4.64	1000	0.006	2.22	23.14	0.045	1.35
GRL	18-20/174	99-32002/313	18-38/395	99-30858/354	GAAC	0.06	0.15	-0.09	0.34	5.65	1000	0.007	2.15	18.83	0.175	0.76
NET	16-50/196	99-28788/300	99-32061/304	99-32121/242	TGGG	0.01	0.05	-0.05	0.10	5.15	1000	0.008	2.10	24.11	0.034	1.47
NET	19-56/140	16-50/196	99-28788/300	99-32121/242	ATAA	0.18	0.06	0.12	3.45	6.78	1000	0.009	2.05	19.07	0.240	0.62
NET	19-14/241	16-50/196	99-32061/304	99-32121/242	TTGG	0.00	0.06	-0.05	0.08	5.70	1000	0.009	2.05	27.52	0.053	1.28
5HTR7	99-28106/185	99-28173/395	99-32181/192	99-32193/258	ACCG	0.27	0.38	-0.11	0.60	2.65	1000	0.012	1.92	14.10	0.023	1.64
5HTR7	99-28171/458	99-28173/395	99-32181/192	99-32193/258	ACCG	0.27	0.38	-0.11	0.61	2.56	1000	0.012	1.92	10.14	0.067	1.17
NET	19-14/241	99-28788/300	99-32061/304	99-32121/242	TGGG	0.01	0.05	-0.04	0.15	3.82	1000	0.012	1.92	25.41	0.025	1.60

846 Total Haplotypes

TABLE 17B

3-Locus Combination

					EM Methods						Permutation Test			Omnibus LR		
Gene	Marker1	Marker2	Marker3	Marker4	Haplotype	Case	Control	difference	Odds Ratio	Chi-Square	Number	p-value	-log p-value	LR statistic	p-value	-log p-value
NET	99-28788/300	99-32061/304	99-32121/242	0	GGG	0.01	0.07	-0.07	0.08	7.92	100000	0.0002	3.77	22.06	0.002	2.65
WFS1	19-19/174	19-17/188	19-18/310	0	CAG	0.01	0.07	-0.06	0.11	6.51	100000	0.0002	3.68	25.36	0.000	3.55
NET	16-50/196	99-32061/304	99-32121/242	0	TGG	0.00	0.07	-0.07	0.06	8.34	100000	0.0008	3.11	21.73	0.012	1.91
NET	19-14/241	19-28/136	16-50/196	0	TCT	0.30	0.43	-0.14	0.55	4.53	100000	0.0009	3.05	20.54	0.001	3.05
GRL	18-20/174	18-38/395	99-30858/354	0	GAC	0.11	0.23	-0.11	0.44	5.18	100000	0.001	2.98	15.29	0.049	1.31
Gbeta3	19-58/162	19-9/45	20-205/302	0	TTC	0.21	0.35	-0.13	0.51	4.89	10000	0.001	2.96	28.52	0.000	4.00
cyp3a4	12-254/180	10-214/279	10-217/91	0	ATC	0.38	0.52	-0.14	0.57	4.41	1000	0.002	2.70	14.23	0.042	1.38
NET	19-14/241	99-32061/304	99-32121/242	0	TGG	0.01	0.06	-0.05	0.08	6.12	1000	0.002	2.70	20.70	0.015	1.82
GRL	99-32002/313	18-31/178	99-30858/354	0	GTT	0.18	0.09	0.10	0.09	4.09	1000	0.003	2.52	9.36	0.188	0.73
GRL	18-31/178	99-30853/364	99-30858/354	0	TAT	0.19	0.10	0.10	2.24	3.83	1000	0.003	2.52	12.79	0.109	0.96
5HTT	18-186/391	18-194/130	18-242/300	0	GCA	0.02	0.08	-0.06	0.27	4.08	1000	0.003	2.52	18.07	0.028	1.55
5HTT	18-186/391	18-198/252	18-242/300	0	GAA	0.02	0.09	-0.06	0.24	5.11	1000	0.003	2.52	17.54	0.031	1.51
DRD3	8-15/126	8-19/372	99-2409/298	0	GGA	0.11	0.04	0.08	3.28	4.31	100000	0.003	2.46	11.57	0.038	1.42
WFS1	19-19/174	19-17/188	16-145/405	0	TGT	0.01	0.06	-0.04	0.24	3.44	100000	0.004	2.41	25.29	0.001	3.27
NET	19-56/140	19-14/241	19-28/136	0	GTC	0.30	0.41	-0.11	0.61	2.92	1000	0.004	2.40	7.06	0.355	0.45
NET	19-56/140	99-32061/304	99-32121/242	0	GGG	0.01	0.07	-0.07	0.11	6.87	1000	0.004	2.40	24.66	0.004	2.40
5HTR6	99-27199/207	99-27207/117	99-28134/215	0	CCT	0.26	0.14	0.12	2.17	4.83	1000	0.004	2.40	15.16	0.081	1.09
5HTT	18-194/130	18-198/252	18-242/300	0	CAA	0.02	0.08	-0.06	0.27	4.15	1000	0.004	2.40	16.28	0.021	1.68
WFS1	19-19/174	19-16/127	19-17/188	0	CGA	0.01	0.06	-0.05	0.21	4.02	1000	0.005	2.30	23.24	0.001	3.00
WFS1	19-19/174	19-17/188	19-18/310	0	CGA	0.12	0.06	0.06	2.16	2.32	1000	0.005	2.30	25.36	0.000	3.55

869 Total Haplotypes

TABLE 17C

2-Locus Combination

					EM Methods						Permutation Test			Omnibus LR		
Gene	Marker1	Marker2	Marker3	Marker4	Haplotype	Case	Control	difference	Odds Ratio	Chi-Square	number	p-value	-log p-value	LR statistic	p-value	-log p-value
WFS1	19-19/174	19-17/188	0	0	CA	0.02	0.08	-0.06	0.21	5.19	100000	0.0005	3.30	20.22	0.000	3.68
Gbeta3	19-58/162	20-205/302	0	0	TC	0.25	0.38	-0.13	0.53	4.68	100000	0.0009	3.04	21.82	0.000	4.52
DRD3	8-15/126	99-2409/298	0	0	AA	0.23	0.37	-0.13	0.52	4.95	100000	0.0012	2.93	10.17	0.014	1.87
WFS1	19-19/174	19-17/188	0	0	CG	0.41	0.28	0.13	1.79	4.09	100000	0.0016	2.79	20.22	0.0002	3.68
NET	19-14/241	19-28/136	0	0	TC	0.31	0.43	-0.12	0.59	3.46	1000	0.002	2.70	8.47	0.044	1.36
NET	99-32061/304	99-32121/242	0	0	GG	0.01	0.09	-0.08	0.15	7.53	1000	0.002	2.70	19.47	0.003	2.52
Gbeta3	19-9/45	20-205/302	0	0	TC	0.24	0.36	-0.12	0.58	3.58	1000	0.002	2.70	17.27	0.001	3.00
GRL	18-31/178	99-30858/354	0	0	TT	0.20	0.10	0.10	2.31	4.19	1000	0.003	2.52	8.22	0.029	1.54
5HTT	18-186/394	18-242/300	0	0	GA	0.03	0.11	-0.08	0.25	6.36	1000	0.003	2.52	13.34	0.016	1.80
5HTT	18-198/252	18-242/300	0	0	AA	0.02	0.09	-0.06	0.24	4.92	1000	0.003	2.52	10.72	0.021	1.68
CHRNA7	99-28722/90	99-32306/409	0	0	CC	0.07	0.29	-0.22	0.19	19.25	1000	0.003	2.52	9.19	0.042	1.38
DRD3	8-15/126	99-2409/298	0	0	GA	0.13	0.04	0.09	3.58	5.16	100000	0.003	2.49	10.17	0.014	1.87
WFS1	19-19/174	19-17/188	0	0	TG	0.01	0.06	-0.04	0.24	3.46	100000	0.004	2.45	20.22	0.000	3.68
5HTR7	99-32181/192	99-32193/258	0	0	TG	0.42	0.28	0.14	1.87	4.16	1000	0.004	2.40	9.27	0.015	1.82
GRL	18-38/395	99-30858/354	0	0	AC	0.44	0.58	-0.14	0.58	4.11	1000	0.004	2.40	8.98	0.023	1.64
CRFR1	99-27091/220	99-27550/48	0	0	GG	0.36	0.25	0.11	1.72	3.20	1000	0.004	2.40	9.03	0.022	1.66
cyp3a4	12-254/180	10-214/279	0	0	AT	0.39	0.52	-0.13	0.60	3.69	1000	0.005	2.30	7.10	0.085	1.07
NET	19-56/140	19-14/241	0	0	GT	0.35	0.46	-0.10	0.65	2.43	1000	0.005	2.30	5.05	0.134	0.87
CHRNA7	99-28722/90	99-28730/351	0	0	TG	0.55	0.36	0.19	2.17	8.05	1000	0.005	2.30	3.95	0.214	0.67
cyp3a4	12-254/180	10-217/91	0	0	AC	0.41	0.53	-0.13	0.60	3.57	100000	0.006	2.26	7.40	0.084	1.08

485 Total Haplotype

Table 18A

				Allele 1 Frequencies				Genotype Frequencies								Numbers	
								Cases			Controls						
Gene	Marker	Allele 1	Allele 2	Cases	Controls	Chi Square	p-value	11	12	22	11	12	22	Chi-square	p-value	Cases	Controls
5HTR6	99-27199-207	C	T	0.72	0.70	0.14	0.71	0.51	0.41	0.08	0.48	0.45	0.07	0.36	0.83	140	94
5HTR6	99-27207-117	C	T	0.48	0.37	5.78	0.02	0.24	0.47	0.29	0.13	0.47	0.40	5.79	0.06	139	93
5HTR6	99-28110-75	C	T	0.64	0.52	5.93	0.01	0.41	0.45	0.14	0.27	0.51	0.22	5.91	0.05	139	90
5HTR6	99-28134-215	C	T	0.51	0.63	6.43	0.01	0.26	0.50	0.24	0.42	0.43	0.15	6.58	0.04	136	93
5HTR7	99-28160-285	A	G	0.49	0.51	0.12	0.73	0.27	0.44	0.29	0.24	0.53	0.23	1.96	0.37	140	91
5HTR7	99-28171-458	A	G	0.34	0.38	0.61	0.44	0.13	0.42	0.45	0.17	0.42	0.41	0.61	0.74	136	90
5HTR7	99-28173-395	C	T	0.51	0.55	0.94	0.33	0.27	0.47	0.26	0.27	0.57	0.16	3.46	0.18	137	94
5HTR7	99-32181-192	C	T	0.28	0.38	5.50	0.02	0.09	0.38	0.54	0.13	0.51	0.36	6.25	0.04	136	85
5HTR7	99-32193-258	G	T	0.70	0.65	1.29	0.26	0.53	0.34	0.12	0.41	0.48	0.11	4.30	0.12	137	85
CHRNA7	99-28722-90	C	T	0.27	0.32	1.24	0.27	0.00	0.54	0.46	0.01	0.62	0.37	2.99	0.22	140	94
CHRNA7	99-28730-351	A	G	0.32	0.38	1.86	0.17	0.04	0.55	0.41	0.03	0.69	0.28	4.93	0.09	139	94
CHRNA7	99-32306-409	C	G	0.62	0.67	1.02	0.31	0.26	0.72	0.01	0.34	0.66	0.00	2.77	0.25	137	91
CRFR1	99-27088-246	A	G	0.42	0.38	1.04	0.31	0.20	0.44	0.36	0.15	0.45	0.40	1.12	0.57	138	93
CRFR1	99-27091-220	A	G	0.44	0.49	1.00	0.32	0.19	0.50	0.31	0.24	0.50	0.26	1.04	0.59	137	92
CRFR1	99-27097-83	C	T	0.55	0.53	0.12	0.73	0.34	0.41	0.24	0.26	0.56	0.19	4.40	0.11	140	90
CRFR1	99-27550-48	A	G	0.20	0.26	1.93	0.17	0.04	0.32	0.64	0.07	0.38	0.55	1.92	0.38	138	91
MLR	16-31-738-A23	C	G	0.55	0.53	0.15	0.70	0.28	0.55	0.18	0.28	0.50	0.22	0.75	0.69	137	92
MLR	99-27563-400	A	G	0.46	0.44	0.07	0.79	0.19	0.54	0.28	0.21	0.46	0.33	1.25	0.54	138	89
MLR	99-28732-133	A	G	0.39	0.31	2.94	0.09	0.16	0.47	0.37	0.09	0.44	0.47	2.98	0.23	137	86
MLR	99-28735-56	C	T	0.71	0.70	0.03	0.87	0.50	0.41	0.09	0.50	0.40	0.10	0.14	0.93	136	88
CRFR2	99-27875-185	C	T	0.36	0.40	1.06	0.30	0.16	0.39	0.45	0.16	0.49	0.35	2.76	0.25	140	94
CRFR2	99-27880-176	C	T	0.53	0.56	0.52	0.47	0.30	0.46	0.24	0.32	0.47	0.20	0.55	0.76	139	93
CRFR2	99-28747-371	C	T	0.39	0.44	0.90	0.34	0.16	0.47	0.37	0.23	0.42	0.35	1.88	0.39	138	91
CRFR2	99-28755-206	A	G	0.56	0.59	0.21	0.65	0.32	0.50	0.19	0.37	0.44	0.19	0.77	0.68	139	93

Table 18B

				Allele 1 Frequencies				Genotype Frequencies								Numbers	
								Cases			Controls			Chi-			
Gene	Marker	Allele 1	Allele 2	Cases	Controls	Chi Square	p-value	11	12	22	11	12	22	square	p-value	Cases	Controls
GRL	99-30853-364	A	G	0.56	0.58	0.24	0.62	0.35	0.41	0.24	0.39	0.39	0.23	0.29	0.87	139	93
GRL	99-30858-354	C	T	0.76	0.84	4.10	0.04	0.59	0.34	0.06	0.71	0.27	0.02	4.04	0.13	140	89
GRL	99-28485-198	G	T	0.19	0.20	0.17	0.68	0.04	0.30	0.66	0.01	0.38	0.61	2.75	0.25	134	89
GRL	99-32002-313	A	G	0.48	0.48	0.00	0.97	0.23	0.51	0.26	0.16	0.65	0.19	4.61	0.10	140	94
GRL	18-20-174	A	G	0.76	0.73	0.76	0.38	0.57	0.39	0.04	0.53	0.39	0.07	1.18	0.55	140	94
GRL	18-31-178	C	T	0.30	0.35	1.55	0.21	0.11	0.38	0.51	0.15	0.40	0.45	1.41	0.49	140	94
GRL	18-38-395	A	T	0.59	0.63	0.76	0.38	0.36	0.46	0.18	0.36	0.54	0.10	3.49	0.17	140	94
MAO A-B	18-2-192	G	T	0.70	0.68	0.19	0.66	0.56	0.26	0.17	0.56	0.24	0.20	0.50	0.78	140	93
MAO A-B	16-215-80	C	T	0.67	0.67	0.00	1.00	0.54	0.26	0.20	0.54	0.24	0.21	0.08	0.96	138	90
MAO A-B	99-26921-210	A	G	0.48	0.52	0.52	0.47	0.33	0.30	0.37	0.31	0.41	0.28	3.17	0.20	139	93
MAO A-B	18-133-293	A	C	0.30	0.27	0.55	0.46	0.16	0.29	0.55	0.14	0.26	0.60	0.52	0.77	140	92
NET	99-28788-300	A	G	0.49	0.47	0.25	0.62	0.23	0.51	0.26	0.21	0.53	0.27	0.32	0.85	314	178
NET	99-32061-304	A	G	0.43	0.40	0.32	0.57	0.19	0.47	0.34	0.14	0.52	0.34	1.41	0.50	320	86
NET	99-32121-242	A	G	0.57	0.54	0.78	0.38	0.34	0.47	0.19	0.30	0.48	0.22	0.74	0.69	313	181
NET	99-32148-315	C	G	0.31	0.27	2.08	0.15	0.09	0.46	0.46	0.09	0.36	0.55	4.23	0.12	315	181
NET	19-28-136	A	C	0.68	0.61	4.35	0.04	0.44	0.47	0.09	0.37	0.49	0.14	4.77	0.09	322	183
NET	19-29-303	C	T	0.21	0.19	0.77	0.38	0.05	0.33	0.63	0.02	0.34	0.64	3.52	0.17	326	183
NET	19-46-322	C	T	0.31	0.31	0.01	0.92	0.07	0.48	0.45	0.07	0.47	0.45	0.03	0.99	317	177
NET	99-32065-303	G	T	0.42	0.43	0.04	0.85	0.18	0.49	0.34	0.14	0.57	0.29	3.02	0.22	317	182
NET	99-32131-312	C	T	0.68	0.71	0.62	0.43	0.47	0.42	0.11	0.49	0.43	0.08	1.12	0.57	315	178
TACR1	99-28761-311	A	G	0.25	0.22	0.78	0.38	0.07	0.36	0.56	0.01	0.41	0.57	4.81	0.09	140	94
TACR1	99-28771-86	C	T	0.48	0.52	0.96	0.33	0.24	0.46	0.29	0.23	0.57	0.19	3.65	0.16	140	94
TACR1	99-28791-291	A	G	0.28	0.26	0.31	0.58	0.09	0.39	0.53	0.04	0.43	0.53	1.76	0.42	140	94
TACR1	99-32361-419	G	T	0.53	0.52	0.07	0.79	0.29	0.49	0.23	0.26	0.52	0.22	0.34	0.84	140	94

Table 18B (cont)

DRD3	8-15-126	A	G	0.64	0.70	1.99	0.16	0.43	0.42	0.15	0.47	0.47	0.06	4.10	0.13	140	94		
DRD3	8-19-372	A	G	0.23	0.28	1.48	0.22	0.09	0.29	0.62	0.03	0.50	0.47	11.32	0.004	140	94		
DRD3	99-2409-298	A	G	0.36	0.40	1.06	0.30	0.14	0.44	0.42	0.11	0.60	0.30	5.31	0.07	140	94		
DRD3	99-339-54	C	G	0.55	0.54	0.01	0.93	0.34	0.42	0.24	0.28	0.53	0.19	2.77	0.25	140	94		

Table 18C

				Allele 1 Frequencies				Genotype Frequencies								Numbers	
								Cases			Controls						
Gene	Marker	Allele 1	Allele 2	Cases	Controls	Chi Square	p-value	11	12	22	11	12	22	Chi-square	p-value	Cases	Controls
CYP3A4	12-254-180	A	G	0.43	0.54	5.86	0.02	0.19	0.47	0.34	0.32	0.45	0.23	5.73	0.06	140	94
CYP3A4	10-214-279	C	T	0.14	0.12	0.35	0.55	0.02	0.23	0.75	0.02	0.19	0.79	0.46	0.79	140	94
CYP3A4	10-217-91	C	T	0.89	0.93	2.01	0.16	0.79	0.19	0.02	0.86	0.13	0.01	1.88	0.39	140	94
5HTT	18-194-130	C	T	0.54	0.52	0.18	0.67	0.31	0.44	0.24	0.20	0.63	0.17	7.73	0.02	140	94
5HTT	18-198-252	A	G	0.52	0.49	0.32	0.57	0.29	0.47	0.24	0.20	0.59	0.21	3.19	0.20	140	94
5HTT	18-242-300	A	G	0.46	0.54	2.89	0.09	0.24	0.44	0.32	0.27	0.54	0.19	4.89	0.09	140	94
5HTT	18-186-391	G	T	0.50	0.53	0.41	0.52	0.27	0.45	0.28	0.26	0.54	0.20	2.37	0.31	140	94
DRD2	16-21-228	A	G	0.21	0.16	1.40	0.24	0.05	0.31	0.64	0.04	0.24	0.72	1.65	0.44	139	93
DRD2	16-23-404	A	G	0.55	0.53	0.09	0.76	0.30	0.50	0.20	0.26	0.55	0.19	0.50	0.78	139	93
DRD2	16-24-175	A	C	0.22	0.16	2.07	0.15	0.06	0.32	0.62	0.04	0.24	0.72	2.19	0.34	138	92
DRD2	16-25-286	C	T	0.70	0.63	2.41	0.12	0.49	0.43	0.08	0.38	0.52	0.11	2.87	0.24	135	93
DRD2	16-22-156	C	T	0.46	0.45	0.07	0.79	0.21	0.50	0.29	0.17	0.56	0.27	0.94	0.63	140	93
Gbeta3	19-58-162	C	T	0.68	0.63	2.24	0.13	0.45	0.47	0.09	0.41	0.43	0.15	4.35	0.11	229	182
Gbeta3	19-9-45	C	T	0.69	0.66	0.60	0.44	0.45	0.48	0.07	0.45	0.42	0.12	4.45	0.11	321	185
Gbeta3	18-355-67	C	T	0.47	0.49	0.43	0.51	0.20	0.55	0.26	0.25	0.47	0.27	2.95	0.23	311	173
Gbeta3	18-353-267	C	T	0.78	0.73	3.58	0.06	0.60	0.37	0.03	0.54	0.38	0.08	6.14	0.05	302	180
Gbeta3	18-388-305	A	G	0.65	0.70	2.05	0.15	0.43	0.44	0.13	0.47	0.45	0.08	2.78	0.25	319	174

Table 18C (cont)

WFS1	19-17-188	A	G	0.57	0.66	8.62	0.003	0.32	0.50	0.18	0.43	0.47	0.10	8.92	0.01	318	180
WFS1	19-19-174	C	T	0.40	0.31	7.66	0.006	0.16	0.48	0.36	0.09	0.43	0.47	7.61	0.02	311	173
WFS1	24-243-346	C	T	0.62	0.70	5.29	0.02	0.37	0.50	0.13	0.48	0.43	0.09	5.70	0.06	321	166
WFS1	99-62531-351	C	T	0.59	0.64	2.77	0.10	0.34	0.51	0.16	0.40	0.49	0.11	2.92	0.23	316	177
WFS1	99-54279-152	C	G	0.59	0.56	0.79	0.37	0.33	0.52	0.15	0.32	0.49	0.20	1.63	0.44	306	174
TH	16-115-343	A	C	0.69	0.76	2.67	0.10	0.51	0.37	0.12	0.60	0.32	0.08	2.37	0.31	132	91
TH	18-251-176	C	T	0.58	0.57	0.02	0.87	0.36	0.44	0.20	0.29	0.57	0.14	4.43	0.11	140	94
TH	18-269-44	A	G	0.36	0.38	0.14	0.71	0.11	0.51	0.39	0.15	0.46	0.39	1.09	0.58	140	94
5HTR2c	18-12-191	A	C	0.15	0.14	0.12	0.73	0.66	0.17	0.77	0.03	0.21	0.76	1.84	0.40	137	94

Table 19

OMNIBUS LR Rank of Haplotypes

	GENE	Affected	Control	Marker1	Marker2	Marker3	Marker4	p-value	-log p-value
1	WFS1	288	148	19-17-188	19-19-174	24-243-346	99-62531-351	0.003	2.49
2		277	154	19-17-188	19-19-174	24-243-346	99-54279-152	0.003	2.47
3	WFS1	274	157	19-17-188	19-19-174	99-62531-351	99-54279-152	0.007	2.17
4	5HTT	140	94	18-186-391	18-194-130	18-198-252	18-242-300	0.013	1.89
5	NET	293	79	99-32061-304	99-32121-242	19-28-136	99-32131-312	0.020	1.69
6	WFS1	283	148	19-17-188	24-243-346	99-62531-351	99-54279-152	0.023	1.65
7	NET	290	78	99-32061-304	99-32121-242	19-28-136	19-46-322	0.032	1.50
8	NET	292	73	99-32061-304	19-28-136	19-46-322	99-32131-312	0.035	1.45
9	NET	291	85	99-32061-304	99-32121-242	19-28-136	99-32065-303	0.038	1.42
10	NET	291	76	99-28788-300	99-32061-304	19-28-136	99-32131-312	0.040	1.40

	GENE								
1	WFS1	297	158	19-17-188	19-19-174	24-243-346		0.001	3.00
2	WFS1	282	166	19-17-188	19-19-174	99-54279-152		0.001	2.85
3	WFS1	291	156	19-17-188	24-243-346	99-54279-152		0.002	2.62
4	WFS1	293	167	19-17-188	19-19-174	99-62531-351		0.005	2.34
5	WFS1	302	150	19-17-188	24-243-346	99-62531-351		0.008	2.11
6	NET	302	79	99-32061-304	19-28-136	99-32131-312		0.017	1.76
7	NET	309	170	19-28-136	19-29-303	19-46-322		0.027	1.57
8	WFS1	287	159	19-17-188	99-62531-351	99-54279-152		0.027	1.56
9	NET	300	85	99-32061-304	99-32121-242	19-28-136		0.037	1.44
10	Gbeta3	217	167	19-58-162	19-9-45	18-355-67		0.039	1.41

Table 19 (cont)

	GENE						
1	WFS1	303	178	19-17-188	19-19-174	0.001	2.92
2	WFS1	312	160	19-17-188	24-243-346	0.001	2.92
3	5HTR7	136	82	99-28106-185	99-32181-192	0.007	2.15
4	Gbeta3	226	180	19-58-162	19-9-45	0.008	2.11
5	WFS1	307	169	19-17-188	99-62531-351	0.012	1.94
6	WFS1	296	168	19-17-188	99-54279-152	0.012	1.94
7	DRD3	140	94	8-15-126	99-2409-298	0.014	1.87
8	5HTR7	133	76	99-32181-192	99-32193-258	0.015	1.82
9	5HTT	140	94	18-186-394	18-242-300	0.016	1.80
10	Gbeta3	294	168	18-353-267	18-338-305	0.033	1.48

Table 20A
Rank of Permutation Tests for Indivdual Haplotypes

							EM methods				Permutation Test			Omnibus LR	
	Gene	Marker1	Marker2	Marker3	Marker4	Haplotype	Case	Control	difference	Chi-Square	number	p-value	-log p-value	p-value	-log p-value
1	GRL	18-20-174	99-32002-313	18-31-178	18-38-395	AGCA	0.06	0.17	-0.12	16.39	100000	0.003	2.57	0.360	0.44
2	GRL	18-20-174	18-31-178	18-38-395	99-30858-354	GTAC	0.11	0.22	-0.10	9.13	100000	0.003	2.55	0.045	1.35
3	GRL	18-20-174	18-38-395	99-30853-364	99-30858-354	GAAC	0.03	0.12	-0.09	12.74	1000	0.003	2.52	0.078	1.11
4	GRL	99-32002-313	18-31-178	18-38-395	99-30858-354	GCAC	0.06	0.14	-0.09	9.82	1000	0.003	2.52	0.159	0.80
5	GRL	99-32002-313	18-31-178	18-38-395	99-30858-354	GTAT	0.13	0.05	0.09	9.55	1000	0.003	2.52	0.159	0.80
6	5HTT	18-186-391	18-194-130	18-198-252	18-242-300	GCAA	0.02	0.08	-0.06	8.26	1000	0.003	2.52	0.013	1.89
7	GRL	99-32002-313	18-31-178	99-30853-364	99-30858-354	GTAT	0.18	0.08	0.10	8.40	1000	0.004	2.38	0.314	0.50
8	5HTR7	99-28106-185	99-28171-458	99-28173-395	99-32181-192	AACC	0.27	0.38	-0.11	2.90	1000	0.005	2.30	0.034	1.47
9	NET	99-32121-242	99-32148-315	19-46-322	99-32131-312	AGTT	0.02	0.07	-0.05	13.54	5000	0.006	2.25	0.273	0.56
10	GRL	18-31-178	18-38-395	99-30853-364	99-30858-354	TAAT	0.14	0.05	0.09	9.35	1000	0.006	2.22	0.090	1.05
11	DRD3	8-15-126	8-19-372	99-2409-298	99-339-54	GGAG	0.06	0.01	0.06	9.29	1000	0.006	2.22	0.045	1.35
12	NET	99-32061-304	19-28-136	19-46-322	99-32065-303	GCTT	0.18	0.30	-0.12	11.71	5000	0.007	2.15	0.155	0.81
13	NET	99-28788-300	99-32061-304	99-32121-242	19-28-136	AGAC	0.16	0.26	-0.11	9.89	5000	0.007	2.14	0.073	1.14
14	GRL	18-20-174	99-32002-313	18-38-395	99-30858-354	GAAC	0.06	0.15	-0.09	11.29	1000	0.007	2.15	0.175	0.76
15	NET	99-32061-304	99-32121-242	19-28-136	99-32065-303	GACT	0.18	0.29	-0.11	10.22	5000	0.008	2.11	0.038	1.42
16	NET	99-32061-304	99-32148-315	19-29-303	99-32065-303	ACTG	0.11	0.03	0.08	9.73	5000	0.008	2.11	0.464	0.33
17	NET	99-32061-304	19-28-136	19-29-303	19-46-322	GCTT	0.17	0.29	-0.12	11.84	5000	0.008	2.10	0.186	0.73
18	NET	99-32061-304	99-32121-242	19-28-136	19-46-322	GACT	0.18	0.30	-0.12	11.06	5000	0.009	2.07	0.032	1.50
19	NET	99-32061-304	99-32148-315	19-28-136	19-46-322	GGCT	0.11	0.22	-0.11	12.79	5000	0.009	2.07	0.218	0.66
20	NET	99-28788-300	99-32061-304	99-32148-315	19-28-136	AGGC	0.11	0.21	-0.10	11.33	5000	0.009	2.06	0.140	0.85

Table 20B
Rank of Permutation Tests for Indivdual Haplotypes

					EM methods					Permutation Test			Omnibus LR	
	Gene	Marker1	Marker2	Marker3	Haplotype	Case	Control	difference	Chi-Square	number	p-value	-log p-value	p-value	-log p-value
1		18-20-174	18-38-395	99-30858-354	GAC	0.11	0.23	-0.11	10.35	100000	0.001	2.98	0.05	1.31
2	WFS1	19-17-188	19-19-174	99-62531-351	ACC	0.01	0.05	-0.04	12.09	5000	0.002	2.80	0.00	2.34
3	cyp3a4	12-254-180	10-214-279	10-217-91	ATC	0.38	0.52	-0.14	8.82	1000	0.002	2.70	0.04	1.38
4	WFS1	19-17-188	19-19-174	24-243-346	GCT	0.34	0.24	0.10	9.80	5000	0.003	2.52	0.00	3.00
5	GRL	99-32002-313	18-31-178	99-30858-354	GTT	0.18	0.09	0.10	8.18	1000	0.003	2.52	0.19	0.73
6	GRL	18-31-178	99-30853-364	99-30858-354	TAT	0.19	0.10	0.10	7.65	1000	0.003	2.52	0.11	0.96
7	5HTT	18-186-391	18-194-130	18-242-300	GCA	0.02	0.08	-0.06	8.17	1000	0.003	2.52	0.03	1.55
8	5HTT	18-186-394	18-198-252	18-242-300	GAA	0.02	0.09	-0.06	10.22	1000	0.003	2.52	0.03	1.51
9	DRD3	8-15-126	8-19-372	99-2409-298	GGA	0.11	0.04	0.08	8.62	100000	0.003	2.46	0.04	1.42
10	WFS1	19-17-188	24-243-346	99-54279-152	ATG	0.01	0.04	-0.03	9.90	5000	0.004	2.36	0.00	2.62
11	5HTR6	99-27199-207	99-27207-117	99-28134-215	CCT	0.26	0.14	0.12	9.66	1000	0.004	2.40	0.08	1.09
12	5HTT	18-194-130	18-198-252	18-242-300	CAA	0.02	0.08	-0.06	8.29	1000	0.004	2.40	0.02	1.68
13	NET	99-32061-304	99-32148-315	19-29-303	ACT	0.11	0.03	0.08	10.38	5000	0.006	2.25	0.38	0.42
14	NET	99-32148-315	19-28-136	19-29-303	GCT	0.21	0.30	-0.09	9.23	5000	0.006	2.24	0.42	0.38
15	cyp3a4	12-254-180	10-214-279	10-217-91	GTC	0.48	0.36	0.12	6.41	1000	0.006	2.22	0.04	1.38
16	NET	99-32061-304	19-28-136	19-46-322	GCT	0.17	0.29	-0.12	11.46	5000	0.007	2.18	0.05	1.33
17	5HTR7	99-28171-458	99-28173-395	99-32181-192	ACC	0.27	0.38	-0.11	5.71	1000	0.007	2.15	0.05	1.28
18	GRL	99-32002-313	18-31-178	18-38-395	GCA	0.07	0.18	-0.11	14.17	1000	0.007	2.15	0.25	0.61
19	GRL	18-31-178	18-38-395	99-30858-354	TAT	0.14	0.05	0.09	8.81	1000	0.008	2.10	0.06	1.25
20	CHRNA7	99-28722-90	99-28730-351	99-32306-409	CAC	0.01	0.05	-0.05	10.75	1000	0.008	2.10	0.04	1.46

Table 20C
Rank of Permutation Tests for Indivdual Haplotypes

					EM methods				Permutation Test			Omnibus LR	
	Gene	Marker1	Marker2	Haplotype	Case	Control	difference	Chi-Square	number	p-value	-log p-value	p-value	-log p-value
1	WFS1	19-17-188	24-243-346	GT	0.37	0.26	0.11	12.09	5000	0.0006	3.22	0.001	2.92
2	WFS1	19-17-188	19-19-174	GC	0.38	0.27	0.11	11.49	5000	0.0012	2.92	0.001	2.92
3	DRD3	8-15-126	99-2409-298	AA	0.23	0.37	-0.13	9.91	100000	0.0012	2.93	0.014	1.87
4	WFS1	19-17-188	99-62531-351	AC	0.44	0.55	-0.11	10.05	5000	0.002	2.74	0.012	1.94
5	WFS1	19-17-188	24-243-346	AT	0.02	0.05	-0.03	8.82	5000	0.003	2.52	0.001	2.92
6	GRL	18-31-178	99-30858-354	TT	0.20	0.10	0.10	8.38	1000	0.003	2.52	0.029	1.54
7	5HTT	18-186-394	18-242-300	GA	0.03	0.11	-0.08	12.72	1000	0.003	2.52	0.016	1.80
8	5HTT	18-198-252	18-242-300	AA	0.02	0.09	-0.06	9.84	1000	0.003	2.52	0.021	1.68
9	CHRNA7	99-28722-90	99-32306-409	CC	0.07	0.29	-0.22	19.25	1000	0.003	2.52	0.042	1.38
10	Gbeta3	19-58-162	19-9-45	CT	0.06	0.01	0.04	10.37	5000	0.003	2.49	0.008	2.11
11	DRD3	8-15-126	99-2409-298	GA	0.13	0.04	0.09	10.32	100000	0.003	2.49	0.014	1.87
12	5HTR7	99-32181-192	99-32193-258	TG	0.42	0.28	0.14	8.33	1000	0.004	2.40	0.015	1.82
13	GRL	18-38-395	99-30858-354	AC	0.44	0.58	-0.14	8.22	1000	0.004	2.40	0.023	1.64
14	CRFR1	99-27091-220	99-27550-48	GG	0.36	0.25	0.11	6.39	1000	0.004	2.40	0.022	1.66
15	cyp3a4	12-254-180	10-214-279	AT	0.39	0.52	-0.13	7.37	1000	0.005	2.30	0.085	1.07
16	CHRNA7	99-28722-90	99-28730-351	TG	0.55	0.36	0.19	8.05	1000	0.005	2.30	0.214	0.67
17	cyp3a4	12-254-180	10-217-91	AC	0.41	0.53	-0.13	7.14	100000	0.006	2.26	0.084	1.08
18	NET	99-32061-304	19-28-136	GC	0.17	0.27	-0.11	9.91	5000	0.006	2.22	0.039	1.41
19	5HTT	18-194-133	18-242-300	CA	0.02	0.08	-0.06	9.41	1000	0.006	2.22	0.03	1.51
20	5HTR6	99-27207-117	99-28110-75	TT	0.32	0.44	-0.12	6.64	1000	0.006	2.22	0.06	1.22

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a contiguous span of at least 12 nucleotides of a sequence selected from the group consisting of the sequences described in Table 7 and the complements thereof.

2. The polynucleotide according to claim 1, wherein said span includes a CNS disorder-related biallelic marker in said sequence.

3. An isolated polynucleotide comprising a contiguous span of at least 12 nucleotides of a sequence selected from the group consisting of the sequences described in Table 9 and the complements thereof, wherein said span includes a CNS disorder-related biallelic marker in said sequence with the alternative allele present at said biallelic marker.

4. An isolated polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides of a sequence selected from the group consisting of the sequences described in Table 9 and the complements thereof, wherein said span includes a CNS disorder-related biallelic marker in said sequence with the original allele present at said biallelic marker.

5. An isolated polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides of a sequence selected from the group consisting of the sequences described in Table 10 and the complements thereof, wherein said span includes a CNS disorder-related biallelic marker in said sequence.

6. The polynucleotide according to any one of claims 2 to 5, wherein said contiguous span is 18 to 35 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide.

7. The polynucleotide according to claim 6, wherein said polynucleotide consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide.

8. A polynucleotide for use in a hybridization assay for determining the identity of a nucleotide at a CNS disorder-related biallelic marker.

9. A polynucleotide for use in a sequencing assay for determining the identity of a nucleotide at a CNS disorder-related biallelic marker.

10. A polynucleotide for use in an allele specific amplification assay for determining the identity of a CNS disorder-related biallelic marker.

11. A polynucleotide for use in amplifying a segment of nucleotides comprising a CNS disorder-related biallelic marker.

12. A use according to any one of claims 8 to 11, wherein said polynucleotide is selected from the sequences described in Table 7.

13. A method of genotyping an individual comprising: (a) obtaining a biological sample comprising a nucleic acid from said individual; (b) determining the identity of a polymorphic base at a biallelic marker from said nucleic acid; wherein said biallelic marker is selected from any one biallelic marker of Table 7; wherein the identity of the polymorphic base determines the genotype of the individual at said position.

14. A method according to claim 13, wherein said CNS disorder-related biallelic marker is selected from the biallelic markers described in Table 7.

15. The method according to claim 14, wherein said CNS disorder-related biallelic marker is selected from the group consisting of 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 99-28106-185, 99-30858-354, 18-20-174, 99-32002-313, 18-31-178, 18-38-395, 99-30853-364, 19-56-140, 19-28-136, 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 16-50-196, 8-19-372, 12-254-180, 10-214-279, 10-217-91, 18-194-130, 18-186-391, 18-198-252, 18-242-300, 20-205-302, 19-58-162, 19-9-45, 19-22-74, 19-88-185, 19-18-310, 19-19-174, 19-17-188, 19-16-127, 99-32148-315, 19-46-322, 99-32131-312, 99-32065-303, 19-44-251, 19-29-303, 18-355-67, 18-353-267, 18-338-305, 16-88-185, 24-243-346, 99-62531-351, 99-54279-152, 99-28171-458, 99-28173-395, 18-186-394, 8-15-126, 99-2409-298, 99-28722-90 and 99-32306-409.

16. The method according to claim 14, wherein said CNS disorder-related biallelic marker is selected from the group consisting of 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 19-28-136, 16-50-196, 19-58-162, 19-9-45, 20-205-302, 24-243-346, 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 19-17-188 and 19-19-174.

17. The method according to claim 13, wherein said biological sample is derived from a single subject.

18. The method according to claim 17, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said subject's genome.

19. The method according claim 13, wherein said biological sample is derived from multiple subjects.

20. The method according to claim 13, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said determining step.

21. A method of determining the frequency in a population of an allele of a CNS disorder-related biallelic marker, comprising:

a) genotyping individuals from said population for said biallelic marker according to the method of claim 13; and

b) determining the proportional representation of said biallelic marker in said population.

22. The method according to claim 21, wherein said CNS disorder-related biallelic marker is selected from the biallelic markers described in Table 7.

23. The method according to claim 22, wherein said CNS disorder-related biallelic marker is selected from the group consisting of 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 99-28106-185, 99-30858-354, 18-20-174, 99-32002-313, 18-31-178, 18-38-395, 99-30853-364, 19-56-140, 19-28-136, 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 16-50-196, 8-19-372, 12-254-180, 10-214-279, 10-217-91, 18-194-130, 18-186-391, 18-198-252, 18-242-300, 20-205-302, 19-58-162, 19-9-45, 19-22-74, 19-88-185, 19-18-310, 19-19-174, 19-17-188, 19-16-127, 99-32148-315, 19-46-322, 99-32131-312, 99-32065-303, 19-44-251, 19-29-303, 18-355-67, 18-353-267, 18-338-305, 16-88-185, 24-243-346, 99-62531-351, 99-54279-152, 99-28171-458, 99-28173-395, 18-186-394, 8-15-126, 99-2409-298, 99-28722-90 and 99-32306-409.

24. The method according to claim 22, wherein said CNS disorder-related biallelic marker is selected from the group consisting of 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 19-28-136, 16-50-196, 19-58-162, 19-9-45, 20-205-302, 24-243-346, 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 19-17-188 and 19-19-174.

25. The method of detecting an association between an allele and a phenotype, comprising the steps of:

a) determining the frequency of at least one CNS disorder-related biallelic marker allele in a trait positive population according to the method of claim 21;

b) determining the frequency of said CNS disorder-related biallelic marker allele in a control population according to the method of claim 21; and

c) determining whether a statistically significant association exists between said allele and said phenotype.

26. The method of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising:

a) genotyping each individual in said population for at least one CNS disorder-related biallelic marker according to claim 13;

b) genotyping each individual in said population for a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome; and

c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency.

27. The method according to claim 26, wherein said haplotype determination method is selected from the group consisting of asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark method, or an expectation maximization algorithm.

28. The method according to claim 26, wherein said CNS disorder-related biallelic marker is selected from the biallelic markers described in Table 7.

29. The method according to claim 28, wherein said CNS disorder-related biallelic marker is selected from the group consisting of 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 99-28106-185, 99-30858-354, 18-20-174, 99-32002-313, 18-31-178, 18-38-395, 99-30853-364, 19-56-140, 19-28-136, 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 16-50-196, 8-19-372, 12-254-180, 10-214-279, 10-217-91, 18-194-130, 18-186-391, 18-198-252, 18-242-300, 20-205-302, 19-58-162, 19-9-45, 19-22-74, 19-88-185, 19-18-310, 19-19-174, 19-17-188, 19-16-127, 99-32148-315, 19-46-322, 99-32131-312, 99-32065-303, 19-44-251, 19-29-303, 18-355-67, 18-353-267, 18-338-305, 16-88-185, 24-243-346, 99-62531-351, 99-54279-152, 99-28171-458, 99-28173-395, 18-186-394, 8-15-126, 99-2409-298, 99-28722-90 and 99-32306-409.

30. The method according to claim 28, wherein said CNS disorder-related biallelic marker is selected from the group consisting of 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 19-28-136, 16-50-196, 19-58-162, 19-9-45, 20-205-302, 24-243-346, 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 19-17-188 and 19-19-174.

31. The method according to claim 28, wherein said haplotype comprises one of the following sets of biallelic markers:

99-28106-185, and 99-32181-192;
18-20-174, 99-32002-313, 18-31-178, and 18-38-395;
18-20-174, 18-31-178, 18-38-395, and 99-30858-354;
18-20-174, 18-38-395, 99-30853-364, and 99-30858-354;
99-32002-313, 18-31-178, 18-38-395, and 99-30858-354;
18-20-174, 18-38-395, and 99-30858-354;
99-32002-313, 18-31-178, and 99-30858-354;
18-31-178, 99-30853-364, and 99-30858-354;
19-56-140, 99-28788-300, 99-32061-304, and 99-32121-242;
19-28-136, 99-28788-300, 99-32061-304, and 99-32121-242;
19-14-241, 19-28-136, 99-32061-304, and 99-32121-242;
19-56-140, 19-28-136, 99-32061-304, and 99-32121-242;
19-14-241, 19-28-136, and 16-50-196;
19-28-136, 99-32061-304, and 99-32121-242;
99-28788-300, 99-32061-304, and 99-32121-242;
99-32061-304, and 99-32121-242;
19-56-140, 19-14-241, 19-28-136, and 16-50-196;
99-28788-300, 99-32061-304, and 99-32121-242;
16-50-196, 99-32061-304, and 99-32121-242;
19-14-241, 19-28-136, and 16-50-196;
19-14-241, 99-32061-304, and 99-32121-242;
19-14-241, and 19-28-136;
99-32061-304, and 99-32121-242;
8-15/126, and 99-2409/298;
12-254/180, 10-214/279, and 10-217/91;
18-186/391, 18-194/130, and 18-242/300;
18-186/394, 18-198/252, and 18-242/300;
19-58-162, 19-9-45, 19-22-74, and 20-205-302;
19-58-162, 19-9-45, 19-88-185, and 20-205-302;
19-58-162, 19-9-45, and 20-205-302;

19-58-162, 19-22-74, and 20-205-302;
19-58-162, and 20-205-302;
19-9-45, and 20-205-302;
19-22-74, and 20-205-302;
19-19-174, 19-17-188, and 19-18-310;
19-19-174, 19-16-127, and 19-17-188;
19-19-174, and 19-17-188; and
19-17-188, 19-19-174, 24-243-346, and 99-62531-351;
19-17-188, 19-19-174, 24-243-346, and 99-54279-152;
19-17-188, 19-19-174, 99-62531-351, and 99-54279-152;
99-32002-313, 18-31-178, 99-30853-364, and 99-30858-354;
99-32121-242, 99-32148-315, 19-46-322, and 99-32131-312;
99-28106-185, 99-28171-458, 99-28173-395, and 99-32181-192;
18-186-391, 18-194-130, 18-198-252, and 18-242-300.
19-17-188, 19-19-174, and 24-243-346;
19-17-188, 19-19-174, and 99-54279-152;
19-17-188, 24-243-346, and 99-54279-152;
19-17-188, 19-19-174, and 99-62531-351;
19-17-188, 24-243-346, and 99-62531-351;
99-32061-304, 19-28-136, and 99-32131-312;
12-254-180, 10-214-279, and 10-217-91;
18-186-391, 18-194-130, and 18-242-300;
18-186-394, 18-198-252, and 18-242-300;
8-15-126, 8-19-372, and 99-2409-298.
19-17-188, and 19-19-174;
19-17-188, and 24-243-346;
19-17-188, and 99-62531-351;
19-17-188, and 99-54279-152;
19-58-162, and 19-9-45;
8-15-126, and 99-2409-298;
18-31-178, and 99-30858-354;
18-186-394, and 18-242-300;
18-198-252, and 18-242-300; and
99-28722-90, and 99-32306-409.

32. The method according to claim 28, wherein said CNS disorder-related biallelic marker comprises one of the following sets of biallelic markers:

99-28788-300, 99-32061-304, and 99-32121-242;
19-14-241, 19-28-136, and 16-50-196;
19-17-188, 19-19-174, and 24-243-176; and
19-58-162, 19-9-45, and 20-205-302.

33. The method of detecting an association between a haplotype and a phenotype, comprising the steps of:

a) estimating the frequency of at least one haplotype in a trait positive population according to the method of claim 26;

b) estimating the frequency of said haplotype in a control population according to the method of claim 26; and

c) determining whether a statistically significant association exists between said haplotype and said phenotype.

34. The method according to either claim 25 or 33, wherein said control population is a trait negative population.

35. The method according to either claim 25 or 33, wherein said case control population is a random population.

36. The method according to claim 56, wherein said CNS disorder-related biallelic marker is selected from the group consisting of 99-27207-117, 99-28110-75, 99-28134-215, 99-32181-192, 99-28106-185, 99-30858-354, 18-20-174, 99-32002-313, 18-31-178, 18-38-395, 99-30853-364, 19-56-140, 19-28-136, 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 16-50-196, 8-19-372, 12-254-180, 10-214-279, 10-217-91, 18-194-130, 18-186-391, 18-198-252, 18-242-300, 20-205-302, 19-58-162, 19-9-45, 19-22-74, 19-88-185, 19-18-310, 19-19-174, 19-17-188, and 19-16-127.

37. The method according to claim 33, wherein said CNS disorder-related biallelic marker is selected from the group consisting of 99-28788-300, 99-32061-304, 99-32121-242, 19-14-241, 19-28-136, 16-50-196, 19-58-162, 19-9-45, and 20-205-302.

38. The method according to claim 33, wherein said haplotype comprises one of the following sets of biallelic markers:

99-28106-185, and 99-32181-192;
18-20-174, 99-32002-313, 18-31-178, and 18-38-395;
18-20-174, 18-31-178, 18-38-395, and 99-30858-354;

18-20-174, 18-38-395, 99-30853-364, and 99-30858-354;
99-32002-313, 18-31-178, 18-38-395, and 99-30858-354;
18-20-174, 18-38-395, and 99-30858-354;
99-32002-313, 18-31-178, and 99-30858-354;
18-31-178, 99-30853-364, and 99-30858-354;
19-56-140, 99-28788-300, 99-32061-304, and 99-32121-242;
19-28-136, 99-28788-300, 99-32061-304, and 99-32121-242;
19-14-241, 19-28-136, 99-32061-304, and 99-32121-242;
19-56-140, 19-28-136, 99-32061-304, and 99-32121-242;
19-14-241, 19-28-136, and 16-50-196;
19-28-136, 99-32061-304, and 99-32121-242;
99-28788-300, 99-32061-304, and 99-32121-242;
99-32061-304, and 99-32121-242;
19-56-140, 19-14-241, 19-28-136, and 16-50-196;
99-28788-300, 99-32061-304, and 99-32121-242;
16-50-196, 99-32061-304, and 99-32121-242;
19-14-241, 19-28-136, and 16-50-196;
19-14-241, 99-32061-304, and 99-32121-242;
19-14-241, and 19-28-136;
99-32061-304, and 99-32121-242;
8-15/126, and 99-2409/298;
12-254/180, 10-214/279, and 10-217/91;
18-186/391, 18-194/130, and 18-242/300;
18-186/394, 18-198/252, and 18-242/300;
19-58-162, 19-9-45, 19-22-74, and 20-205-302;
19-58-162, 19-9-45, 19-88-185, and 20-205-302;
19-58-162, 19-9-45, and 20-205-302;
19-58-162, 19-22-74, and 20-205-302;
19-58-162, and 20-205-302;
19-9-45, and 20-205-302;
19-22-74, and 20-205-302;
19-19-174, 19-17-188, and 19-18-310;
19-19-174, 19-16-127, and 19-17-188;
19-19-174, and 19-17-188; and
19-17-188, 19-19-174, 24-243-346, and 99-62531-351;
19-17-188, 19-19-174, 24-243-346, and 99-54279-152;
19-17-188, 19-19-174, 99-62531-351, and 99-54279-152;

99-32002-313, 18-31-178, 99-30853-364, and 99-30858-354;
99-32121-242, 99-32148-315, 19-46-322, and 99-32131-312;
99-28106-185, 99-28171-458, 99-28173-395, and 99-32181-192;
18-186-391, 18-194-130, 18-198-252, and 18-242-300.
19-17-188, 19-19-174, and 24-243-346;
19-17-188, 19-19-174, and 99-54279-152;
19-17-188, 24-243-346, and 99-54279-152;
19-17-188, 19-19-174, and 99-62531-351;
19-17-188, 24-243-346, and 99-62531-351;
99-32061-304, 19-28-136, and 99-32131-312;
12-254-180, 10-214-279, and 10-217-91;
18-186-391, 18-194-130, and 18-242-300;
18-186-394, 18-198-252, and 18-242-300;
8-15-126, 8-19-372, and 99-2409-298.
19-17-188, and 19-19-174;
19-17-188, and 24-243-346;
19-17-188, and 99-62531-351;
19-17-188, and 99-54279-152;
19-58-162, and 19-9-45;
8-15-126, and 99-2409-298;
18-31-178, and 99-30858-354;
18-186-394, and 18-242-300;
18-198-252, and 18-242-300; and
99-28722-90, and 99-32306-409.

39. The method according to claim 33, wherein said haplotype comprises one of the following sets of biallelic markers:

99-28788-300, 99-32061-304, and 99-32121-242;
19-14-241, 19-28-136, and 16-50-196;
19-17-188, 19-19-174, and 24-243-176; and
19-58-162, 19-9-45, and 20-205-302.

40. The method according to either claim 25 or 33, wherein said phenotype is a CNS disorder.

41. The method according to either claim 25 or 33, wherein said phenotype is a response to an agent acting on a CNS disorder.

42. The method according to either claim 25 or 33, wherein said phenotype is a side effect to an agent acting on a CNS disorder.

43. The method according to claim 25, wherein the identity of the nucleotides at all of the biallelic markers described in Table 7 is determined in steps a) and b).

44. The method of administering a drug or treatment comprising:

a) obtaining a nucleic acid sample from an individual;

b) determining the identity of the polymorphic base of at least one CNS disorder-related biallelic marker according to the method of claim 13 which is associated with a positive response to said drug or treatment, or at least one CNS disorder-related marker which is associated with a negative response to said drug or treatment; and

c) administering said drug or treatment to said individual if said nucleic acid sample contains at least one biallelic marker associated with a positive response to said drug or treatment, or if said nucleic acid sample lacks at least one bialleic marker associated with a negative response to said drug or treatment.

45. The method of selecting an individual for inclusion in a clinical trial of a drug or treatment comprising:

a) obtaining a nucleic acid sample from an individual;

b) determining the identity of the polymorphic base of at least one CNS disorder-related biallelic marker according to the method of claim 13 which is associated with a positive response to said drug or treatment, or at least one biallelic marker associated with a negative response to said drug or treatment in said nucleic acid sample; and

c) including said individual in said clinical trial if said nucleic acid sample contains at least one biallelic marker which is associated with a positive response to said drug or treatment, or if said nucleic acid sample lacks at least one biallelic marker associated with a negative response to said drug or treatment.

46. The diagnostic kit comprising a polynucleotide according to any one of claims 2, 3, 4 and 5.

47. The use of a polynucleotide in a hybridization assay for determining the identity of a nucleotide at a CNS disorder-related biallelic marker.

48. The use of a polynucleotide in a sequencing assay for determining the identity of a nucleotide at a CNS disorder-related biallelic marker.

49. The use of a polynucleotide in an allele specific amplification assay for determining the identity of a CNS disorder-related biallelic marker.

50. The use of a polynucleotide in amplifying a segment of nucleotides comprising a CNS disorder-related biallelic marker.